Transit Access Modeling Improvements for Application in Transit Assignment Models

A Thesis
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Benjamin David Nault-Maurer

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Alireza Khani, Adviser

June 2017

Acknowledgements

Achievements such as advanced degrees do not happen through the work of one person. Any achievement is a compilation of the work of many, and therefore much credit needs to be given.

I would first and foremost like to thank my advisor, Professor Alireza Khani. It has truly been my pleasure to work with you during my studies. Your compassion, kindness, wisdom, and patience have been recognized and appreciated more times than you know. I am grateful to have such a close and reliable mentor.

I would also like to acknowledge the members of the University of Minnesota Transit Lab. We have become an amazing team, and I am truly grateful to have such wonderful and dear friends.

On another, more professional note, I would like to acknowledge Metro Transit for providing, through the Minnesota Geospatial Commons, much of the data used for validation. I would also like to thank the University of Texas at Austin for allowing me to use the VisTool with FAST-TrIPs to visualize route and path outputs.

Dedication

To my parents and siblings,
who have unconditionally supported and encouraged me to achieve my goals.

This work would not be possible without the unwavering support and understanding of my beloved family, of whom I hold dearest.

Abstract

Transit ridership is sensitive to the amount of time riders need to walk to access transit, as a small increase in walking time can have a large impact on the path utility. Therefore, an improvement in the access distance calculation should be able to provide overall improvements to a transit assignment model. The current state of practice for modeling transit access by walking uses straight line connectors from the center of transportation analysis zones, which were designed for highway demand modeling and may be too large of an area to accurately predict walking distances. Therefore, a new method is presented which models walking access along a network by using land units (e.g. blocks, parcels) without increasing computational complexity. Also, by calculating distances over a network, more realistic distances are calculated, and obstacles such as rivers or freeways are appropriately accounted for. This method was also applied to park and ride lots to assess ridership modeling improvements on a more holistic scale. This research presents a case study where a schedule based transit assignment model implements the proposed strategy on the Twin Cities network. While significant improvements in ridership accuracy were not found on a system level for walking access, lower level analysis (neighborhood) showed consistent improvements where aggregation data such as population and employment were homogeneous and present. For park and ride access, inconsistencies between scenarios were observed and the results of the application of the general access model to a more specific assignment is realized. Therefore, as each scenario performed best in different analysis, it is suggested to use the access link scenario that is most appropriate for the desired analysis. Mainly, for large scale, system level analysis the larger land use units (e.g. TAZ) can provide reasonable assignment, but for lower level analysis, such as stop level boardings or route load profiles, a more detailed scenario, such as presented in this research, should be used.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Access to transit is a key aspect of a transit trip. The majority of transit riders access and egress transit by walking, and typically only use transit if a stop is within a reasonable walking distance from their home or place of work. This prevents transit from being a viable option for some, but it also poses a problem for modeling transit riders. Namely, how do we accurately account for this sensitive and important access measure?

There are many studies on measuring transit accessibility in the transportation planning literature. Usually transit access is measured through high resolution GIS networks, and a transit accessible area is determined for planning purposes [1]. In the most basic method, an accessible area is determined by a buffer around routes or stops, and any point inside the buffer is assumed accessible to transit. In some cases, access to transit is defined as a function of distance, and a decreasing pattern is observed in the number of trips taken by transit when the distance to the transit stops increases [2]. In other types of accessibility analysis, access to major destinations is evaluated from an origin point in the network [3], and sometimes at different times of the day [4]. In most cases, a simplified transit path is used to calculate an accessibility measure to key destinations.

While transit accessibility is studied from the planning perspective to analyze transit coverage, equity, safety, etc., the connection with network models is not well-established. In other words, most of the assignment models in the literature do not use an accurate access model to estimate passengers' walking times. This can lead to inaccurate model predictions as the access to transit is a small but highly influential aspect of a transit trip. In transit assignment modeling, this sensitivity is measured in terms of perceived travel time. This means if one minute of riding in a bus is perceived as one minute of travel time, riders may perceive one minute of walking as the equivalent to two or more minutes of riding on the bus [5]. Therefore, small inaccuracies in walking distances are exaggerated when viewed over the course of a trip.

In this research, a method is proposed that utilizes land use data to model walking access to transit stops, which can then be integrated with a schedule-based transit assignment model. That is, a transit access sub-model is developed and is integrated with a schedule-based transit assignment model.

Transit assignment models usually simplify walking links by using straight lines from Transportation Analysis Zone (TAZ) centroids to transit stops, which are usually located on nodes in the highway network. This approach leads to a couple shortcomings. Firstly, a straight line is not a good representation of a walking path. Secondly, using zone centroids as the transit demand generation points is an unrealistic assumption. This simplification may be appropriate for modeling auto access to major roads as the TAZs were designed to be used in highway demand models. However, from the user behavior perspective, walking time is less desirable than transit in-vehicle time, and a small change in walking time may affect the whole path significantly. Therefore, the importance of walking links in the transit assignment model was the motivation for this thesis, with the intention to integrate a more sophisticated access model with an advanced transit assignment model.

1.2 Motivation and Problem Statement

Creating accurate measures of walking access is difficult for many reasons. Firstly, calculating an accurate walking distance is difficult due to how sensitive and integral it is to transit ridership. This calculation of access is based on a distance, where a transit stop is less accessible the further the distance is. Currently this distance is calculated using a buffer, or straight line, distance from the center of a TAZ to a transit stop. This is a great simplification as people cannot travel in a straight line from their home to a transit stop. They must travel on sidewalks, roads, and paths, which is better known as the network. Measuring distance on a network, the path riders must travel on, provides a great improvement by itself.

Secondly, TAZs were designed for use in traffic assignment and are relatively large for use in walking access, with the average size of a TAZ in the study area being 3.4 mi^2. Driving for an extra 0.5 - 1 miles is a reasonable assumption for traffic assignment, but walking that distance could make or break a rider's decision to use transit, especially if that distance is measured in a straight line, and is more like 2 - 3 miles on the network.

Improving the access model from a buffer distance to a network distance is relatively straightforward. However, assuming all the population is at the center of the TAZ is a large assumption. If the population of a TAZ is focused on one side of the TAZ, the transit stops on the other side of the TAZ could be 2+ miles away from anyone living in the TAZ, but only 1 mile from the center. Therefore, a transit stop could be deemed accessible when in reality it is not.

To address this problem, this research aims to break the TAZ assignment into smaller, more realistic walking assignments (i.e. blocks, parcels). However, this isn't straight forward as there is not origin destination (OD) demand data at these smaller land units. Even if there were OD demand data at this level, breaking the network into these smaller units would increase computational complexity significantly. Therefore, this research has developed a model to take access distances from those smaller land units and apply their distances to the overall TAZ distance. This way the transit assignment can provide the detail of the smaller unit's accuracy while keeping the data format of the TAZ and maintaining the assignment's complexity.

The objective of this research is to improve the overall accuracy of transit assignment models by improving the accuracy of access link calculations. These access link calculations are improved through the development of a model that addresses the buffer distance calculation, acknowledges the bias of using the large TAZ regions for transit assignment, and implements a new accessibility measure. This research therefore provides a more accurate measure of access distance in a format that is general for all assignment models and doesn't increase assignment complexity.

1.3 Contributions and Thesis Organization

Two contributions can be considered in this research: one being the state-of-art and one being the state-of-practice. First, recent developments both in accessibility modeling and in GIS-based data analysis experiments have been combined to estimate the accessibility to transit stops from each TAZ. In this step, data such as block- and parcel-based land use data, General Transit Feed Specification data [6], and an open-source street network, which was used to replace the buffer area with a network distance from transit stops, have been taken advantage of. In this research the new method is evaluated by testing five different combinations of input data and comparing them to the current method on various analysis levels.

The second and main contribution of this research is to integrate the accessibility model with a schedule-based transit assignment model to improve the network-level travel forecasting capability. In other words, after analyzing the accessibility to each transit stop in the first part of the research, the estimated walking links are used as a part of the transit network in the assignment model. This is a major contribution in the field since none of the accessibility models have gone so far as to be used in a regional transit assignment model for travel-forecasting purposes. On the other hand, the current transit assignment models in practice do not model access links using a true accessibility model. The experiments in a real case study involving the Twin Cities show improvements in the assignment model's ability to estimate ridership on a route, neighborhood, stop, and origin destination (OD) path level analysis.

This thesis is divided into 7 chapters. After this introduction, a literature review is presented in Chapter 2 that discusses the problem and literature in more detail. Following the literature review, Chapter 3 discusses the data required and used for this research. Chapter 4 discusses the methodology and provides a proof of concept of the new accessibility calculation method. Next, Chapters 5 and 6 discuss the application of this research in a real case study involving Minneapolis, MN and Saint Paul, MN. The new accessibility measure is applied to both walking access and park and ride access in Chapters 5 and 6 respectively. Finally, Chapter 7 concludes the thesis, providing key findings, discussion, and future research suggestions.

Chapter 2: Literature Review

2.1 Walking Access

In many studies, transit access is assumed within a buffered area around a transit route or stop [1] [2] [7] [8] [9] [10] [11] [12] [13] [14] [15]. The buffer-based approaches choose a threshold distance to represent the coverage area [9]. Grava 2003 [16] suggests that a transit stop is attractive for an individual living within 0.25 miles (400 m), measured as a straight-line distance.

This distance has come under great scrutiny as many studies have performed research on origin destination surveys and have found little empirical evidence to support the 0.25 mile threshold [9] [17] [18] [19]. This research has focused on determining the actual distance riders walk to transit based on factors such as trip purpose, land use, and density. El-Geneidy et. al. [17] looked at a case study involving an origin destination (OD) survey from Montreal, Canada and compared walking access distance based on the mode of transit used (i.e. bus, LRT). This study found the 0.25-mile access distance actually tended to underestimate access, and that people tended to walk longer distances to reach commuter and rail services. It suggests to use a custom access distance based on stop/station type and land use. Larsen [18] also looked at an OD survey from Montreal, and also determined the 0.25-mile distance underestimated access to transit stops. Larsen focused on developing appropriate distances based on trip purpose, trip mode (walking or biking) and socio-economic characteristics. However, other studies from Badland et. al. [20] and Horner and Murray [21] found the 0.25-mile buffer to overestimate walking distance. These articles all view the 0.25-mile buffer critically, and note that walking distance should be estimated for local conditions as the distances travelled to access transit varies significantly between regions.

Whether the 0.25-mile distance over or under-estimates the access distance, this threshold is not precise and must depend on the network attributes such as network connectivity, demographic information, the percentage of elderly people, grade, etc. [22].

As an alternative, a distance decay function may be used. Distance decay functions look at the frequency of trips based on the distance riders walked. This is an important estimation, as it has been shown that pedestrians who walk to transit first look to minimize the distance and time of the walking aspect of their trip [5] [23]. A decay function is typical in all modes of transportation and represents the fact that, for a given mode, the frequency of trips decreases by increasing distance. Several studies can be found that estimate a decay function for different transportation models [3] [10]. This distance decay function is often integrated into buffer based distances by using different rings or bands of accessibility and weighting the rings based on distance [9] [24] [25].

One of the major benefits of using the buffer distance is the ease of implementation. Using Geographic Information System (GIS) technology, planners and engineers can easily determine the threshold distance and look at land use characteristics near transit stations or transit lines. However, the main limitation of using the buffer based distance for accessibility is its inability to account for physical obstacles such as rivers or freeways. Therefore, buffers could determine a stop that is on the other side of a river is accessible to riders with no way to access that stop.

The deficiencies of a buffer-based methodology have encouraged researchers to explore other approaches using additional sources of information. As one of the first studies, O'Neill et. al. 1992 [26] proposed the network ratio method, which defines the catchment area of each transit stop as the ratio of the total length of streets within walking distance of 0.25 miles to the total length of all the streets in the network. Other studies, such as Anderson and Landex [9] have used a similar approach where they used a detour factor to reduce buffer distance without calculating network distance. They noted that although the detour factor helped, it was not precise enough to provide accurate data, especially as their network to buffer distance ratios ranged from 0.37 – 0.76. Anderson and Landex [9] also looked at adding a time resistance factor which would account for elevations' effect on walking distance.

As the buffer distance mainly suffers from lack of detail, much of the more recent literature involves comparisons between buffer and network distances. Zhao et. al. 2003

[1] used network distance in a high-resolution street network to improve accessibility models and to incorporate natural and man-made barriers. A comparison between the buffer, network ratio, and parcel-network methods is presented in Biba et. al. 2010 [7]. The results show the buffer and network ratio methods generate higher estimates of access than the parcel-network method.

These studies also bring the concept of land use into the accessibility calculations. Early on, Peng and Dueker [27] looked at a buffer distance around the transit lines, and looked at land use data on the census block group level. Models have continued to develop, and Zhao 1998 [15] showed the advantage of using land use data to better predict the population distribution that is accessible to transit. By using land use to model accessibility, more detailed and flexible assignments can be done. However, there are difficulties of using land use characteristics. A main problem is that many times the required data is in too aggregate a form to provide useful information [28]. For example, population and demographic information could be on a smaller scale than origin destination data. Without a proper way to aggregate this information, this disaggregate data and the detail it can provide cannot be used. This research aims to make the more disaggregate data and detail available on a more aggregated scale. This added detail should be able to make current assignment models more flexible and accurate without altering the assignment model or adding complexity.

2.2 Park and Ride Access and Integration

After walking, driving is typically the second highest access mode for public transit [29] [30]. Out of those trips, a high majority access transit through park and ride facilities. These lots are often located in more suburban areas, as they provide convenient access to transit for low density areas where transit is not a feasible option. These facilities allow riders to drive and park their vehicles at a location in order to access transit. As these riders are choice riders (they drive to transit instead of driving to their destination), park and ride users represent a unique transit rider profile. The majority of these trips include regular commuters, and these trips tend to be much longer than the typical transit rider's [30].

Typically, an express route is connected to these lots that go into a downtown or other highly congested area, which encourages drivers to use transit instead of driving a personal vehicle. However, as park and ride facilities appeal mainly to choice riders, it can be difficult to yield maximum utilization, and there is a great importance placed on park and ride location.

Much of the literature on park and ride facilities considers how these lots play into the accessibility of the transit system as a whole. Many of these studies use surveys to understand rider's behavior [28] [30]. These studies emphasize the importance of accessibility in transit networks, and the utilization of these facilities. Many conclude that riders are more affluent, regular commuters who have longer transit trips to downtown areas who would like better security at these facilities [28] [30]. As these studies mainly come from a planning perspective, they focus on utilization and future development rather than park and ride accessibility or transit assignment and ridership models.

Of those studies that do focus on modeling park and ride access, there are a few special characteristics of modeling park and ride access that aren't apparent in modeling walking access. Although the use of land use characteristics is consistent for calculating possible demand, many riders who use park and rides have the inherent option to drive. This means park and ride users have a transfer built into their transit trip. Transfers are typically seen as costly for a rider, and difficult to model appropriately. Therefore, careful attention must be taken in order to handle this access type. Studies such as Hendricks and Outlander [31] have worked on improving a regional forecasting model by including intermediate stop choices in a person's choice set. This could mean intermediate personal stops, or a park and ride facility. By including these options in a more general travel forecasting model (as opposed to a transit specific model), it allowed them to more accurately account for passengers to choose to take transit through park and ride facilities.

Another important characteristic of park and ride access is the similarity between auto and transit utilities. As driving is much faster than walking, many drivers will not drive in a direction opposite their destination to access transit, as they could more easily drive to their destination. As riders are much less likely to drive in a direction opposite their destination, especially when compared to riders accessing transit by walking, a backtracking feature is often used. By integrating a back-tracking measure and a maximum distance cutoff (either by straight line or network distance) such as in Farhan and Murray [32], a somewhat parabolic access area is generated, with the vertex pointing toward the downtown, urban area and expanding radially outward. This is an important and unique trait of park and ride modeling that comes about due to the relative speed and utility of auto vehicles as compared to transit. Even though these backtracking features are unique and important to park and ride access, this research focuses on the application of the proposed methodology to walking and park and ride access and its improvement in transit ridership estimation. Therefore, it does not take a back-tracking or other park and ride specific feature into account.

2.3 Distinction from Literature

The main distinction of this research from other literature is the implementation of transit accessibility to improve ridership estimation. Other studies mentioned above mainly evaluate the level of accessibility of transit facilities to measure the effectiveness of a public transit system in providing service. However, in this research, accessibility is blended with transit assignment to model users' behavior, experienced travel time, and ridership for a current network condition. In comparison to studies that use buffer distances to strictly measure accessibility, this study uses a distance of 1 mile as the maximum walking distance and 10 miles as the maximum driving distance to a transit stop and park and ride lot respectively. This is only a cutoff distance however, as a stop is considered accessible if it is within the distance, and then the distance from a transportation analysis zone (TAZ) or block to that stop or lot is calculated separately. In some scenarios these distances are even aggregated to provide a distance that is most likely to be travelled. This allows the transit assignment model to predict which route a rider will take based on the distance they must walk or drive to access transit, and whether or not that distance is reasonable when considering their entire trip. This is unique as the studies above mainly look at accessibility for planning purposes rather than assignment purposes, and many assignment models do not accurately calculate this sensitive access distance.

The difficulty in applying transit accessibility to transit assignment model access links is that transit assignment models need a general format of access links. Currently, models require distances from a TAZ to a transit stop. Therefore, this paper presents a method where transit stop accessibility is integrated into transit access calculations in a way that is generalized for input into transit assignment models.

Chapter 3: Data

3.1 Land Use

The transportation analysis zone (TAZ) data was provided by the local governing agency, the Metropolitan Council of the Twin Cities. The Twin Cities region consists of 3030 TAZs, which contain 77,552 blocks and over 1.07M parcels. Table 1 provides a breakdown of the land use characteristics, and it can be seen that an average TAZ can be broken into 28 blocks and 421 parcels. This shows the possible improvement in the level of detail that is gained by assigning passengers at the block and parcel level.

Block and parcel data came from the United States Census Bureau. Block population was used because although population on the TAZ level was available, it was not consistent between the block and TAZ datasets. Also, LEHD employment data [33] was available on the block level. Therefore, the sum of population and employment was chosen as the main indicator for transit demand at the block level and was used as the weighting factor in the formulation. While population and employment were available on a block level, they were not available on the parcel level.

Table 1: Land Use Characteristics

Number of TAZs	3,030
Number of Blocks	77,552
Number of Parcels	1,073,077
Average Blocks per Zone	28
Average Parcels per Zone	421
Average Population per Zone	1,160
Average Population and Employment per Zone	1,740

3.2 Roadway Network

OpenStreetMap (OSM) [34] is an open source street map that is editable by any user and is updated constantly. The data from OSM was found to be more detailed than data provided by government sources, especially in high density areas. The OSM data was especially suited for this project's use as the enhanced detail was largely in the form of small, residential streets where people are more likely to be walking. Although this detail was well suited for this research, the detail added great complexity to the network, as can be seen in Table 2.

This network was adjusted so only the roads that were usable by foot traffic were analyzed. This meant removing road categories such as motorways, construction, raceway, trunk highways, freeways, and highways from the available network. This alone is an improvement over the buffer method, as pedestrians are not allowed to walk along or cross obstacles such as freeways, rail lines, or rivers.

Table 2: Road Network Properties

	Coarse Network	OSM Network (raw)	OSM Network (walking)
Number of Nodes	167,117	1,301,376	1,205,150
Number of Links	370,420	1,413,808	1,305,945
Total Miles	39,748	172,417,825	157,426,732

3.3 Transit Network (GTFS)

General Transit Feed Specification (GTFS) [6] data was provided by the local transit agency Metro Transit and used for the locations of the TAZs and transit stops for the transit assignment model runs. GTFS is a powerful tool for transit agencies, researchers, and application developers as it provides detailed and consistently updated information about a transit networks' schedule, stop locations, routes, route times, stop times, transfer locations, fares, and much more [35]. GTFS is maintained by Google, and has feeds that are constantly updated by transit agencies. Table 3 provides a breakdown of the GTFS files and their contents.

The network used in this study was from 2011 as this was near to the time the ridership survey was conducted and the land use data was created. This network included 13,891 stops and 191 different bus and rail transit routes. A more detailed breakdown of the transit network can be found in Table 4. As the GTFS data is updated at irregular times, problems with matching the network with validation data were encountered, as discussed in Chapter 5.

Table 3: GTFS File Structure (Source: Google Transit APIs [35])

GTFS File	Contents
stop.txt	stop_id, stop_code, stop_name, stop_desc, stop_lat, stop_lon, zone_id, location_type, parent_station
transfer.txt	from_stop_id, to_stop_id, transfer_type, min_transfer_time
shape.txt	shape_id, shape_pt_lat, shape_pt_long, shape_pt_sequence, shape_dist_traveled
stoptime.txt	trip_id, arrival_time, departure_time, stop_id, stop_sequence, stop_headsign, pickup_type, drop_off_type, shape_dist_traveled
fare_attribute.txt	fare_id, price, currency_type, payment_method, transfers, transfer_duration
agency.txt	agency_id, agency_name, agency_url, agency_timezone, agency_lang, agency_phone
fare_rule.txt	fare_id, route_id, origin_id, destination_id, contains_id
route.txt	route_id, agency_id, route_short_name, route_long_name, route_desc, route_type, route_url, route_color, route_text_color
trip.txt	route_id, service_id, trip_id, trip_headsign, trip_shortname, direction_id, block_id, shape_id
frequency.txt	trip_id, start_time, end_time, headway_secs
calendar_dates.txt	service_id, date_ exception_type
calendar.txt	service_id, monday, tuesday, wednesday, thursday, friday, saturday, sunday, start_date, end_date

Table 4: Transit Network Characteristics

Total Number of Stops	13,891
Number of Bus stops	13,857
Number of LRT Stops	27
Number of Commuter Rail Stops	7
Number of Bus Routes	188
Number of LRT Routes	2
Number of Commuter Rail Routes	1

3.4 Transit Ridership

The model outputs of this study were compared to data from Metro Transit to validate its accuracy. Both a Travel Behavior Inventory On Board Survey from 2010 [29] as well as automated passenger count (APC) [36] data for transit services from 2013 were provided by Metro Transit and used for validation. This data was useful in calibrating and validating the model's accuracy. Any discrepancies between results from the assignment on the 2011 transit network to the 2013 APC data are discussed in Chapter 5. The APC data is also a random sample for routes, as only 85% of the buses have APC devices [36]. These buses are randomly distributed throughout the routes, and only take a sample of the actual trips taken.

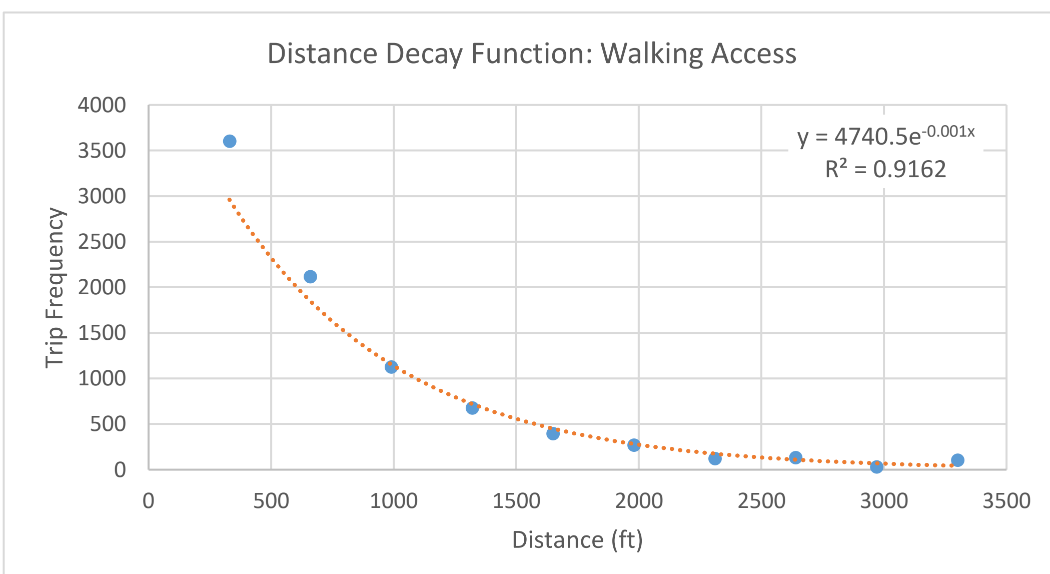

Figure 1: Distance Decay Function for Walking Access Trips in the Twin Cities

The 2010 Travel Behavior Inventory On Board Survey was another integral dataset for this project. The demand for the transit assignment model was created from this dataset and was considered fixed in this research. The survey was also used as a measure of actual ridership, namely for path analysis in Section 5.9. Even though this study didn't use a distance decay function, Figure 1 depicts the distance decay function as measured using the On Board Survey. This shows that most riders who access transit by walking are within 0.2 miles (4 minutes' walk) of a transit stop, and the ridership decreases dramatically after 0.2 miles. This information is useful for comparing the model results to the real data.

Table 5 provides a detailed summary of the breakdown of how riders access transit in the Twin Cities. From this table, it can be seen that most (79%) riders access transit by walking. This is evidence of how crucial the walking access measurement is to accurate transit ridership estimation. Also, the second highest access mode is by driving, namely through park and ride services. This was motivation for the second part of this research where this access measurement is applied to park and ride access links.

Finally, Table 6 shows the frequency of ridership by walking access. It can be seen most riders (56%) take transit 5 or more times per week, which could indicate most riders are using transit for commuting. This makes the transit assignment more reliable as the assignment is for a typical weekday.

Table 5: Access Mode Distribution (based on the 2010 On Board Survey)

Row Labels	**Number of Surveys**	**Ridership**	
Walk	15,747	219,677	79%
Drive	4,241	30,450	11%
Bike	304	4,289	2%
Shared Ride	788	8,632	3%
Dropped off	239	1,722	1%
Other	616	9,333	3%
Missing Value	414	4,846	2%
Grand Total	**22,349**	**278,950**	**100%**

Table 6: Frequency of Walking Trips (based on the 2010 On Board Survey)

Row Labels	**Average of Blocks Walked**	**Ridership**	
This is the First Time	3.01	1,603	1%
1-4 Days Per Month	2.72	6,120	3%
2-4 Days Per Week	2.50	37,422	17%
5+ Days Per Week	2.42	121,927	56%
A Few Times Per Year	3.36	3,716	2%
Missing Value	2.58	8,564	4%
Not available		40,325	18%
Grand Total	**2.48**	**219,677**	**100%**

Chapter 4: Methodology

4.1 Motivation of Concept

Walking access is an important calculation when it comes to modeling transit systems due to the sensitivity walking has on path choice. A slight increase in walking time can have a large impact on how attractive a path may be. Therefore, the current practice of calculating a straight-line distance from a TAZ centroid to a transit stop may not provide the sensitivity to accurately calculate walking access links. Therefore, a lower level stop accessibility calculation may provide more accurate results.

However, traditionally calculating the access links from a block level, with n blocks for example, would dramatically increase the complexity of transit assignment by a factor of $O(n^2)$ by creating n^2 origin and destination (OD) pairs. In the Twin Cities, this would mean going from 9.2M OD pairs to over 6.0B OD pairs. Therefore, by calculating the walking access distance from each block to a stop and taking a population and employment based average to aggregate back to a TAZ level, an effective distance can be calculated that represents the most likely distance a person in the TAZ would need to walk to access the stop without sacrificing complexity. By using the effective walking distance, there is no need to generate walking links or connectors from each block, which would significantly increase the size of the network, and the complexity remains unchanged. This improves the access link's accuracy without increasing the transit assignment's complexity, and this method can be applied to any land use unit (e.g. blocks, parcels).

To test the new methodology, a transit assignment model called FAST-TrIPs was used [12]. FAST-TrIPs is an open source, schedule based transit assignment and simulation program that models individual route choice and user experiences. By taking a person's origin, destination, and transit preferences, FAST-TrIPs assigns a person to a vehicle trip if transit is accessible during their designated departure and arrival time. This preference is determined through a nested logit model that considers rider's preference and value of walking time, waiting time, transfer time, and other factors.

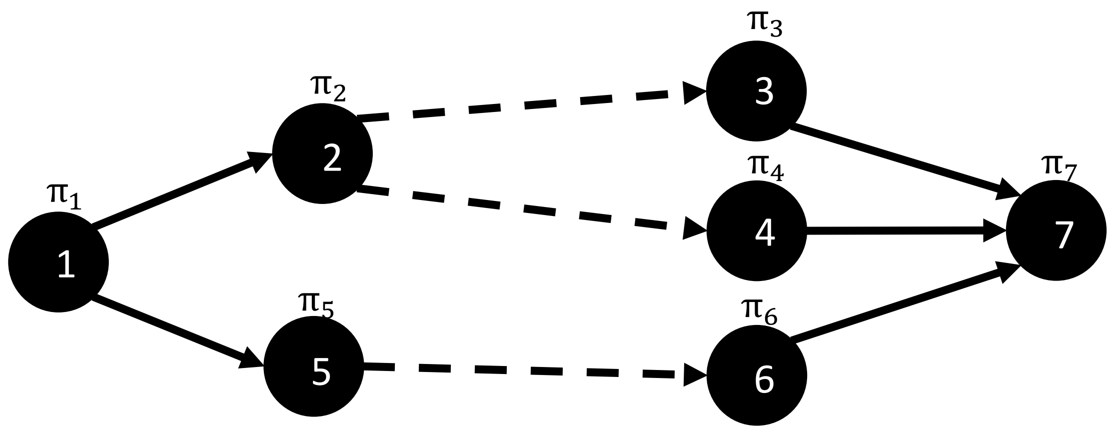

Figure 2: Example of Transit Network

FAST-TrIPs utilizes hyperpath theory to assign passengers to paths. Hyperpaths were introduced by Nyugen and Pallottino in 1988 [37], and have been prevalent in transit and traffic assignment ever since. Hyperpaths provide a way to assign passengers to a transit network based on the probability that any given path is used by a rider, based on their preferences.

To understand the importance and power of hyperpaths, let's look at an example using Figure 2, where β_i is the probability a node i is traversed, and $a_{i,j}$ is the conditional probability link i, j is traversed given node i is traversed. Passengers starting at node 1 and going to node 7 have two initial choices: node 2 or node 5. The probability of going to either node is calculated based on the probability of traversing the link to that node, namely $\beta_2 = \beta_1 a_{1,2}$ and $\beta_5 = \beta_1 a_{1,5}$. If they now move to node 2, they then have the choice to go to node 3 or node 4. The probability of traveling to node 3 then becomes $\beta_3 = \beta_2 a_{2,3}$, and so on until they reach the destination at node 7. More generally, on the path from node r to s, the probability β of traveling through some node j is given by

$$\beta_j = \sum_{i \in B_j} \beta_i \, a_{i,j} \tag{1}$$

where B_j is the backwardstar or the list of nodes leading up to node j, and $\beta_r = \beta_s = 1$.

The concept of hyperpaths is at the basis of the new formulation, as the transit assignment FAST-TrIPs uses hyperpaths to assign passengers to the network. In a hyperpath the probability of choosing an initial walking link is determined by the entire path's link probabilities. However, in the formulation below, the probability a rider chooses a walking path is determined solely on the distance it takes to get to a stop, not the entire transit trip. In this case, the entire trip isn't a concern because the transit assignment uses hyperpath theory to determine a passenger's chosen trip, where the walking links calculated here are one link in the trip, and one input into the model. Therefore, an effective walking distance can be calculated without worrying about future links as the transit assignment takes care of that recursively when the model is run.

4.2 Formulation of Concept

Given zones $i \in Z$, stops $j \in S$, and blocks $k \in K$ with population π_k, the current practice uses the distance d_{ij} from a TAZ centroid to a stop as the walking distance. With embedded logit route choice model used in the transit assignment model, the probability that passengers from zone i choose stop j among accessible stops to zone i is

$$P_i(j) = \frac{e^{-\theta(d_{ij} + u_j)}}{e^{-\theta(d_{ij'} + u_{j'})}} \quad (2)$$

where u_j is the combined utilities of stop j in connecting zone i to a destination zone, and θ is the logit dispersion factor. Given the transit demand from zone i, the number of passengers choosing stop j is

$$x_i(j) = P_i(j) \sum_{k \in K_i} \pi_k \quad (3)$$

Applying the same concept to blocks (if OD demand in block level was available), the probability of stop j being chosen by passengers in block k will be

$$P_k(j) = \frac{e^{-\theta(d_{kj} + u_j)}}{e^{-\theta(d_{kj'} + u_{j'})}} \quad (4)$$

and therefore, the demand from zone i to stop j will be

$$x_i(j) = \sum_{k \in K_i} \pi_k P_k(j) \quad (5)$$

Intuitively, (3) and (5) should be equal, and the effective value for walking distances, $\bar{d}_{ij}$, will be determined which represents the micro-level behavior of demand originating from smaller units (i.e. blocks). This effective distance will be

$$\bar{d}_{ij} = -\frac{1}{\theta} ln\left(\frac{\sum_{k\in K_i} \pi_k P_k(j)}{\sum_{k\in K_i} \pi_k}\right) + C_i \tag{6}$$

or

$$\bar{d}_{ij} = -\frac{1}{\theta} ln\left(\frac{\sum_{k\in K_i} \pi_k \frac{e^{-\theta d_{kj}}}{\sum_{j'} e^{-\theta d_{kj'}}}}{\sum_{k\in K_i} \pi_k}\right) + C_i \tag{7}$$

where C_i is a constant that can be determined by evaluating the formula at one instance of the network. However, because the logit model only uses the difference in the utility of alternatives, and C_i is canceled out when calculating the probability of each stop, there is no need for calculation of C_i for the assignment model.

4.3 Example of Concept

To show a simple example and proof of the concept, imagine a TAZ that contains two nearby stops and four blocks, as depicted in Figure 3. Of the two stops, stop 1 is closer when measuring in a straight line, and stop 2 is further away. However, it can be seen that each block has a certain population, and the population is mainly around the base of the TAZ in blocks 3 and 4, which is closer to stop 2.

This weighting of the population near the base of the TAZ shows one of the strengths of this new method, which is that this type of population distribution exists in the real world but isn't taken into account in the TAZ straight scenario. However, in this example it will be shown how the new method actually shifts passenger assignment from the TAZ to the stop that is closer to the location of the majority of the population of the TAZ. This is more of a realistic assignment, as riders tend to minimize walking distance and therefore will go to stops that are closer to their location rather than the center of the TAZ.

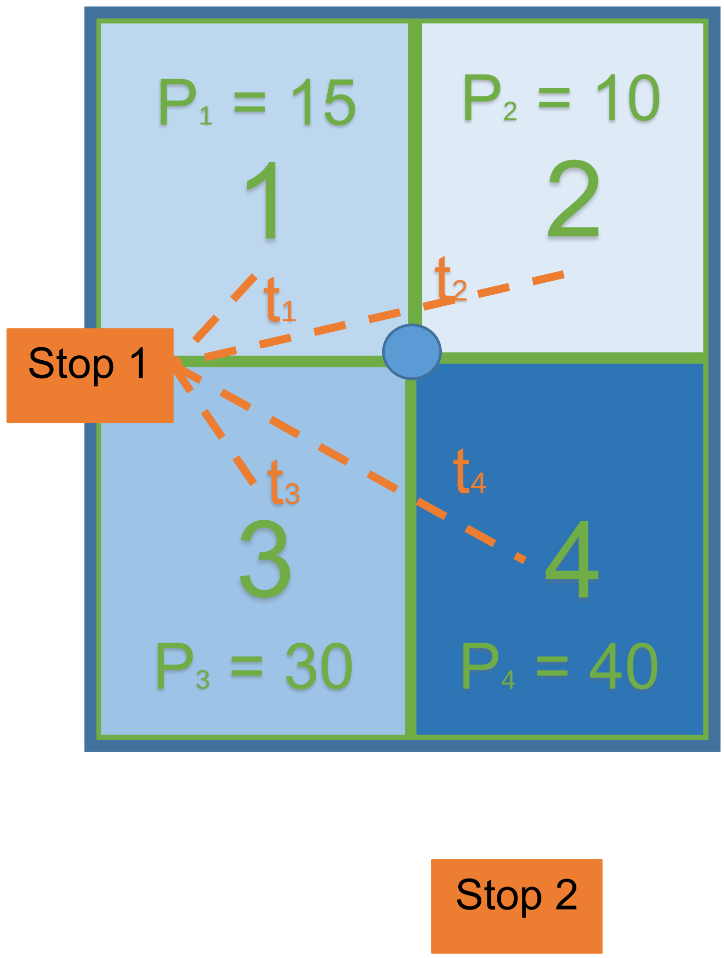

Figure 3: Example Transportation Analysis Zone (TAZ)

For this example, 3 scenarios are presented. Firstly, the TAZ straight scenario where distances for the entire TAZ are estimated as the straight line distance from the center of the TAZ to the stop location. This is the current state of practice, and the logit model and assignment are represented in Equations 2 and 3 respectively. Secondly, the block OD scenario where passengers are assigned based on each block's straight line distance from the center of the block to the stop. This scenario is what is depicted in Equations 4 and 5. This assignment cannot simply be performed as this would dramatically increase the computational cost of the assignment. Even if this could be computed with an OD matrix with this level of detail, OD data at the block level does not exist. Finally, the proposed concept where distances from each block are calculated and the effective distances from the entire TAZ to that stop is determined, as described by Equation 6 and

7. Again, this is an arbitrary distance that represents the typical distance one person in the TAZ would have to walk to the stop. The total number of assigned passengers in all three of these scenarios should be the same, as the transit demand is fixed for this model and therefore passengers cannot be created nor removed from transit ridership in the assignment.

Given the distances provided in Table 7 the disutility parameter associated with walking distance can be calculated. In this case, the disutilities were estimated as the total walking time from the center of the block to the stop, given a person walked at 3 mph (20 min/mile). The transit assignment is sensitive to this disutility measure, and this is a separate source for model calibration in addition to improving walking access links. This disutility plays a role in the hyperpath calculation, as explained above, with regards to the transit assignment, of which the walking access distances plays a role. However, for this simple example, having the disutility be equal to the total walking time works just fine.

Once the disutilities are calculated, the logit model is used to calculate the probability a person from the TAZ or block will choose either stop. Based on that probability and population, and the scenario, the passengers can be assigned to the transit stops. The actual transit assignment the model would use these walking distances as one part of the hyperpath calculated for assignment. Here, probabilities are calculated independent of future links.

Based on the results provided in Table 7, the same number of passengers are assigned to each stop in the case where all blocks are used as origin locations and the case using the new distance calculation method. This is the most significant result, as it shows the same assignment in the scenario where the average distances are weighted by population and where the blocks are treated individually, but the estimation is on the TAZ level. This shows that the new method accurately improves the assignment on the TAZ level as if it were on the block level. Also, passengers are assigned to the stop that is closest to the highest population, instead of the distance from the center of the TAZ.

Table 7: Example Distances, Travel Times, Probabilities, and Assigned Passengers

Stop	TAZ Straight	Average Block (block 1)	Average Block (block 2)	Average Block (block 3)	Average Block (block 4)	Effective Distance
Walking Distances (mi)						
1	0.28	0.1	0.4	0.4	0.1	0.028
2	0.40	0.6	0.6	0.3	0.1	0.042
Disutility (travel times $t_{k,s}$)						
1	5.50	2.00	8.00	2.00	10.00	0.56
2	8.00	12.00	12.00	6.00	2.00	0.43
Probabilities						
1	0.92	1.00	0.98	0.98	0.00	0.57
2	0.08	0.00	0.02	0.02	1.00	0.43
Passengers Assigned						
1	88	54				54
2	7	41				41
Total	95	95				95

4.4 Implementation of Concept

To implement this concept on a real network, each stop within one mile (either straight or network) of the center of a block was considered accessible. As discussed, ridership decreases dramatically after about 0.2 miles, but in this step of assignment, only all possible links are generated, so the distance can be more than 0.2 miles. Once a block-stop pair was considered accessible, the distance between them was calculated. Then, for each TAZ, distances from each block inside the TAZ to each stop accessible to that block were calculated and aggregated together using the proposed formula. In this study, a sum of the population and the total employment for each block was used to weight averages (i.e. as π_k). This provided a more realistic depiction of transit demand than using population alone, or not using population at all. Using this method and the aggregation concept explained above, the effective walking distance for a potential transit rider was calculated.

By recalculating the distance from each block inside a TAZ to a stop accessible to any block in the TAZ, it is possible that effective walking distances can be quite large. However, this is still reasonable as the transit assignment model uses a logit choice model to determine if a transit trip is accessible to a potential rider based on the rider's preferences. Therefore, if a rider needs to walk an unreasonable distance, that rider may choose to not take that transit route.

Chapter 5: Improvements in Walking Access

5.1 Transit Assignment Model

This study uses the FAST-TrIPs schedule based transit assignment model [38]. FAST-TrIPs uses a logit-based hyperpath for simulating passengers' route-finding and user experiences [39]. This model uses the GTFS data as discussed in Section 3.3, although it only uses the calendar.txt, routes.txt, trips.txt, stop_times.txt, stops.txt, and shapes.txt files.

During the time of writing, FAST-TrIPs was being calibrated to fit the Twin Cities region. One part of this is improving walking access calculations, but these walking access links are only one part of a transit trip. Another key factor for transit passenger assignment is route choice. Based on a report on route choice from the local government agency, The Metropolitan Council of the Twin Cities, the parameters in Table 8 represent the route choice parameters used in the model [40]. This model doesn't show the walking links to be as sensitive as discussed earlier, but these parameters vary based on location.

Table 8: Route Choice Parameters for Twin Cities Region

Parameter	Perceived Travel Time (min)
In-Vehicle Time	1.00
Waiting Time	2.76
Access / Egress Walking Time	0.82
Transfer Penalty	7.50

From this model, passenger measures such as route choice and travel time can be calculated. From a system perspective, measures such as load, boardings, alightings, number of transfers, passenger trip profiles, and more can also be calculated. This data can then be compared to similar data obtained from automated passenger count (APC) data for calibration and validation. The APC data is collected using sensors located at the vehicle's doorways, allowing them to measure when riders access or egress the vehicle. Based on the vehicle's location when these boardings and alightings take place, this data can be tied back to a stop location. The data used in this research is from the year 2013, and is aggregated to represent the daily boardings and alightings for each route at each stop. This is important as there are discrepancies when data is this aggregated, which will be discussed later. However, this data is extremely helpful for a transit agency as such detailed information can be useful for planning purposes as well as model calibration.

5.2 Access Link Preparation

Initially, this study was intended to be completed in free and open source GIS software, namely QGIS [41], so this method was accessible to all users. Due to the complexity of the network, many routing options and packages were attempted in the QGIS software, but none were time efficient. Therefore, the data was exported from the QGIS software for preparation and network routing, and instead run on a terminal command line using the Python programming language [42]. Once the roadway network was prepared, the transit stop network, TAZ data, block data, and parcel data were all integrated with the network. With the network prepared outside of QGIS software, and using Dijkstra's [43] one to all shortest path algorithm, the code to prepare the walking access link input file ran in about 8 hours with a machine with normal computing capabilities (i5-4590S CPU @ 3.00 GHz and 8.00 GB memory).

5.3 Improvement Measures

In assessing the effectiveness and robustness of the new walking accessibility concept, different scenarios were tested. Comparisons were made between the base scenario that used TAZ straight line distance to 5 scenarios using the methodology introduced above: network distance using TAZs, straight-line distance aggregating at the block and parcel level, and network distance aggregating at the block and parcel level. It is important to reiterate that the parcel level data did not have population data, and therefore an unweighted average distance to the stops was used. Other weighting factors were tested, such as the square footage of a parcel, but no improvements were found. This isn't to say this or other weighting measures couldn't provide improvements. The block level analysis used the population and employment weighted methodology introduced in Section 4.2.

Three measures were used to determine the accuracy of the model's assignment, as in Tavassoli et. al. [14]. The R-Squared (R^2) measure is calculated between the observed ridership from APC data and the scenario outputs to show the degree of accuracy of the model. The Percent Root Mean Squared Error (%RMSE) was also used to determine the closeness of results, which can be seen in Equation 8

$$\%RMSE = 100 * \frac{\sqrt{\frac{\sum_{i=1}^{N}(Modeled_i - \ Observed_i)^2}{N}}}{\frac{\sum_{i=1}^{N} Observed_i}{N}} \tag{8}$$

where $Modeled$ is the transit assignment model outputs, $Observed$ is the APC data, and N is the number of predictions.

Finally, a GEH measure is used to determine goodness-of-fit. GEH measures are often used in transportation models, especially for traffic simulations, as it is a good way to depict the accuracy of a model with a wide variety of volumes. GEH is also used in transit modeling, where it can be used when systems' routes have a wide range of ridership. The GEH measure is depicted in Equation 9.

$$GEH = \sqrt{\frac{2(Modeled_i - \ Observed_i)^2}{(Modeled_i + \ Observed_i)}} \tag{9}$$

Typically, a GEH value less than 5 predicts a good fit, and a traffic model has a good fit if 85% of the GEH values are under 5 [44] [45] [46]. As transit models are more complex than traffic models, typically a transit model shows good fit if 60% of GEH values are under 5 [14] [47].

5.4 Limitations

This study had some limitations due to model calibration and data availability. First, this project was one effort to improve the assignment model's route choice model calibration. Therefore, the level of calibration that would normally be necessary for a base case scenario in a forecasting model is not seen here. Therefore, this study focuses on the relative improvements of the assignment's prediction rather than absolute accuracy.

Secondly, the network used is from 2011, while the APC data is from 2013 as APC data from 2011 or 2012 were not available. This mismatch is more prevalent in analysis where there were changes made to the routes between 2011 - 2013, but this isn't the case for every route. Therefore, only routes that were unchanged during this time were analyzed.

Thirdly, population and employment data were not available on the parcel level. This limited the analysis of the methodology as equal comparisons between the improvements of aggregating distances on a parcel level versus a block or zone level cannot be done.

Fourthly, while the Travel Behavior Inventory 2010 On Board Survey has a good representation of travel patterns in the Twin Cities region, in the path level analysis some scenarios didn't have many data points. An expansion factor was used to aggregate the data up to predicted levels, but some scenarios' detail suffered due to having few sample points to aggregate. Due to this data availability problem, only scenarios that had the most data present were used.

Lastly, the demand for the transit assignment model is generated from the 2010 On Board Survey, with a few assumptions being made. The survey is broken into many departure times, namely early AM, AM peak, midday, PM peak, and late PM. To input into the model, off peak demand was assumed to be a random combination of early AM, midday, and late PM. AM peak and PM peak were randomly distributed and used for the peak demand. The survey also included biking demand, but biking access is not taken into account, so this demand was integrated into the walking demand.

5.5 Access Link Analysis

The differences in the properties of the walking access links can be seen in Figure 4 and Table 9. In most cases, similar values are seen in the 3 major categories of scenarios: base, straight line distance, and network distance. Maybe the most surprising difference is in the number of access links, as the number of access links is nearly halved when compared to the straight line and original cases. This is most likely due to the difference in area covered by 1 mile measured in a straight line versus 1 mile measured on a network, as depicted in Figure 5. This naturally leads to a decrease in number of accessible stops. It can even be seen in Figure 5 the TAZ network scenario doesn't show any of the stops inside the TAZ are accessible. Figure 6 depicts the difference in the accessible areas for a TAZ that borders a river. It can be seen that although both the block straight and TAZ straight scenarios show accessible stops on the other side of the river, the block and TAZ network scenarios prevent this error. This is a good example of this research and the improvements it has made.

Table 9: Walking Access Link Characteristics

Description	Original	TAZ Network	Block Straight	Block Network	Parcel Straight	Parcel Network
Number of Access Links	135,907	46,676	212,837	101,605	223,384	115,728
Number of Accessible Stops	13,883	12,514	13,891	12,490	13,883	13,469
Zones with Accessible Stops	1,575	1,252	1,717	1,597	1,713	1,520
Average Distance from Zone to Stop	0.68	0.67	0.95	1.23	1.03	1.25
Average Distance from Stop to Zone	0.66	0.67	0.92	1.19	0.97	1.24
Average Stops per Zone	86	37	124	64	130	76
Average Zones per Stop	10	4	15	8	16	9

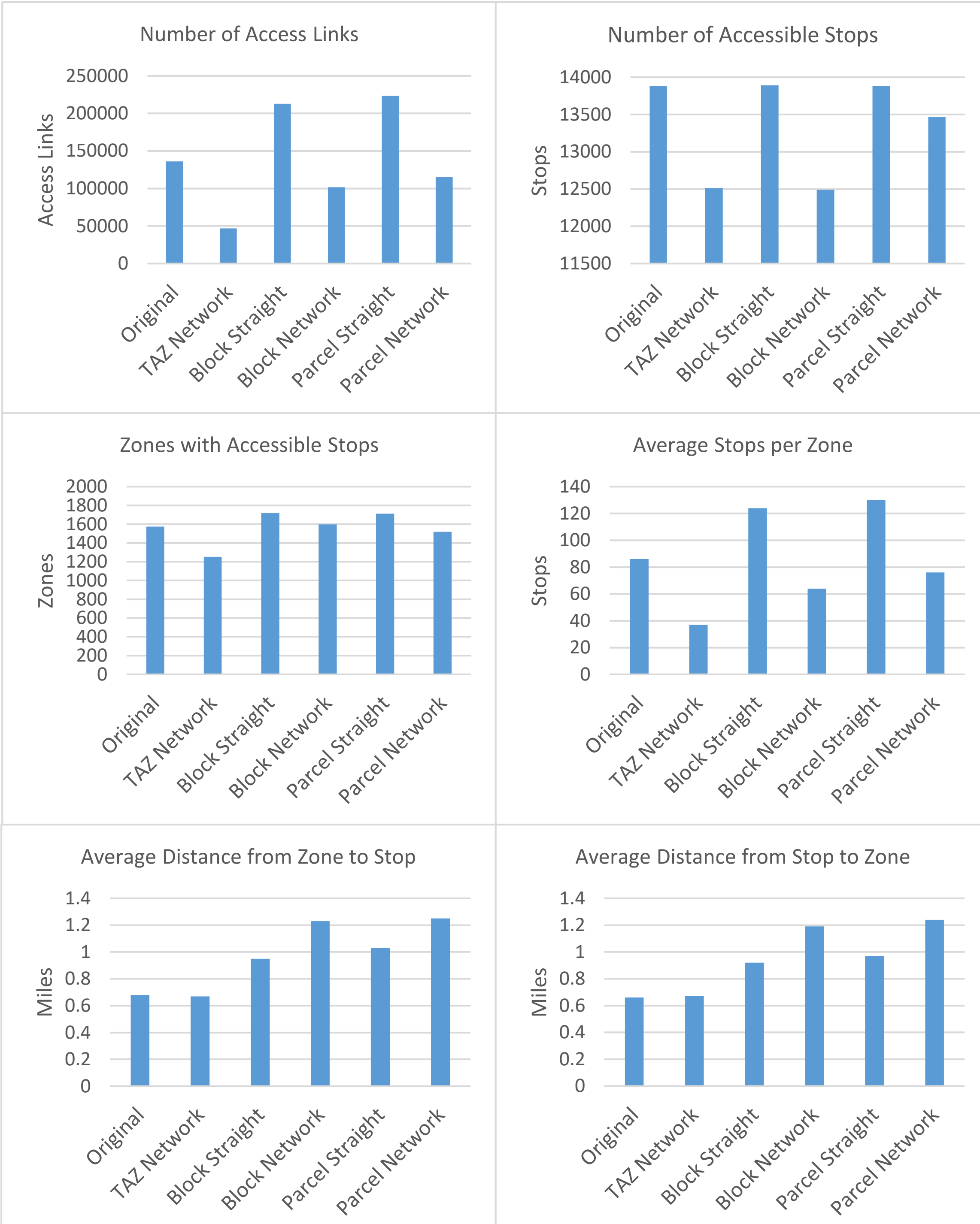

Figure 4: Characteristics of Walking Access Links

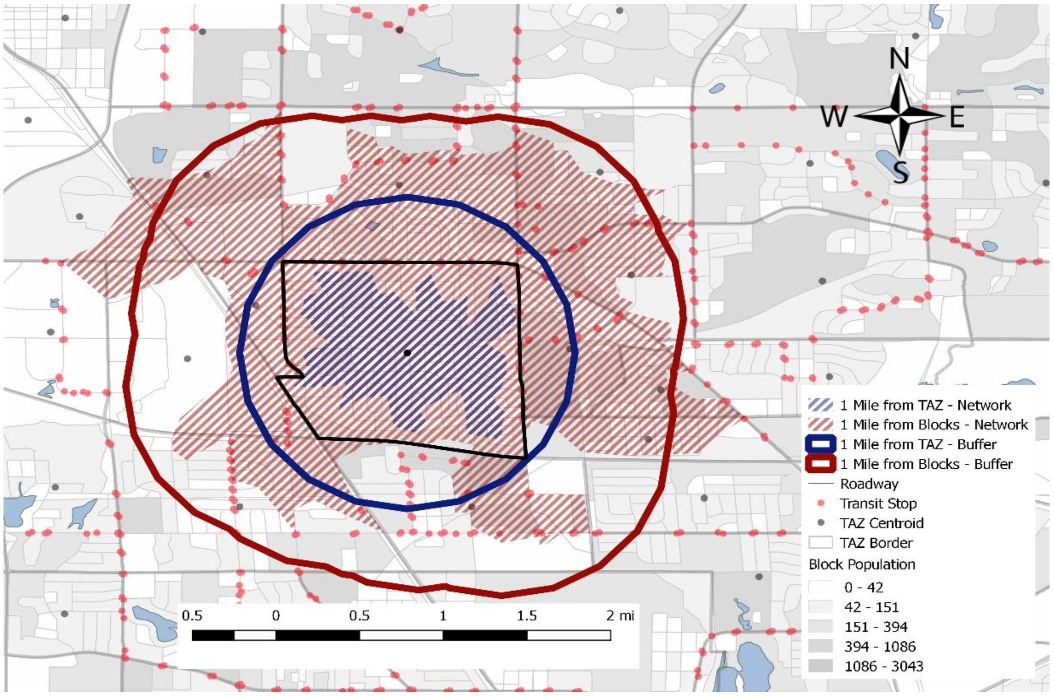

Figure 5: Network and Buffer Distances for TAZ 1084 Center and Block Average

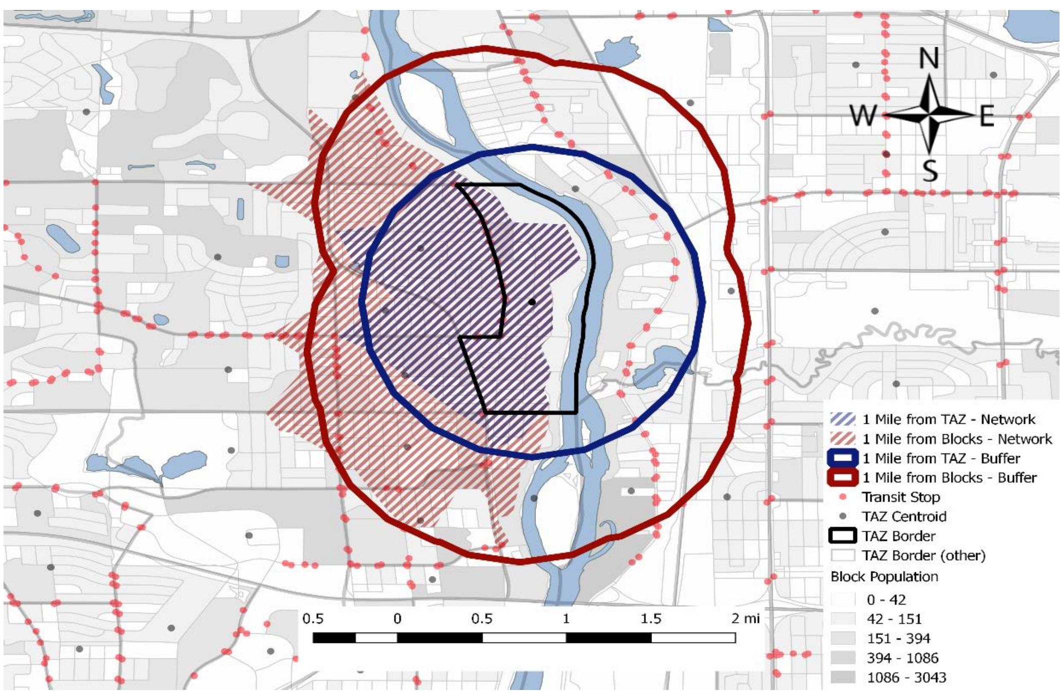

Figure 6: Network and Buffer Distances for TAZ 1076 Center and Block Average

Increases in the average distance from zones to stops and vice versa can also be seen. The block network scenario has nearly 1,500 fewer stops than comparable scenarios. The other 4 scenarios noted 99-100% of the stops in the network were accessible, while the block network scenario showed that 90% were accessible. This could be due to the elimination of unwalkable paths in the network. The difference between the parcel network and block network scenario may disprove this, but the smaller parcel size could account for more accessible paths pedestrians could take. This could also be due to the population weight, as blocks without any measured population do not contribute to the aggregated effective walking distance.

5.6 System Level Analysis

When looking at system level performance, improvements are not seen but accuracy doesn't decrease. Comparisons between scenarios can be seen in Figure 7 and Figure 8, which shows the total ridership by route and stop, respectively, for the six different scenarios. For routes, the original access link file depicts ridership well on a system level, but the straight-line block and parcel scenarios show a decrease in accuracy and fit. For the block network scenario, an increase in accuracy but decrease in fit can be seen. For the parcel network scenario, the same type of improvements are not seen, either in the accuracy or fit. This shows how the new methodology improves modeling rider behavior using the network distance on a system level only for the scenario where both the network distance and population and employment weighting are utilized. An increase in %RMSE can be seen in all scenarios, but the block and parcel straight scenarios have lower values than the block and network scenarios. Similarities in the GEH measure can be seen, but the statistic still does not fall above the desired 60% measure.

In looking at system boardings by stop, the model doesn't fit the stops as well as it fits the routes. Boardings by stop can fluctuate much more than route level ridership. The block network scenario fits the APC ridership the best, with the TAZ network having a slightly less accurate fit. For all the scenarios low %RMSE measures and high %GEH<5 measures are seen, which show the relative error is reasonable, but the fits and R^2 values are not reasonable. This analysis is also difficult to predict accurately on a system level.

When looking at ridership characteristics, such as in Table 10, differences between the scenarios aren't observed. The network scenarios show higher average travel time, possibly due to the increased average walking distance. The number of riders in the system is relatively consistent, although there is a slight drop in ridership in the network scenarios. This is most likely due to the effect on the number of accessible stops due to network distances, as discussed in Section 5.5. However, this doesn't mean the assignment is worse. The total demand is aggregated from the Travel Behavior 2010 On Board Survey, and it is not a steadfast number for ridership. Due to the way the schedule-based transit assignment model assigns passengers, if a transit trip is not accessible or available during the time a rider wishes to leave, that rider will not be assigned to a transit route. Also, the transit assignment model tends to underestimate waiting time as it assumes riders are knowledgeable about the transit schedule, and therefore will arrive at the transit stop at the optimal time to catch their mode of travel. In regards to the average number of transfers, the assignment model tends to under-predict the number of transfers in all scenarios except for the block network. This scenario shows great improvement as the average number of transfers is 0.75, as calculated from the Travel Behavior Inventory 2010 On Board Survey.

Table 10: Ridership Profiles

	Original	TAZ Network	Block Straight	Block Network	Parcel Straight	Parcel Network
Route Ridership %RMSE	96.5	111.6	103.4	105.2	99.3	113.9
Number of Assigned Passengers	213,992	180,687	216,773	207,032	216,690	203,747
Average Travel Time (min)	42.72	47.86	43.55	49.95	43.71	49.1
Average Walking Access Time (min)	7.43	8.04	8.44	9.73	8.6	9.64
Average Number of Transfers	0.55	0.61	0.47	0.74	0.46	0.51
Average Transfer Wait Time (min)	4.84	5.12	4.78	4.75	4.88	4.88
Average Transfer Walk Time (min)	0.83	0.81	0.84	0.84	0.83	0.82
Average In-Vehicle time (min)	18.23	17.46	18.63	17.87	18.77	18.31
Average Egress Time	7.15	8.04	7.77	7.66	9.18	8.83
Total Travel Time (sum)	35.92	37.18	37.48	39.39	39.21	39.72

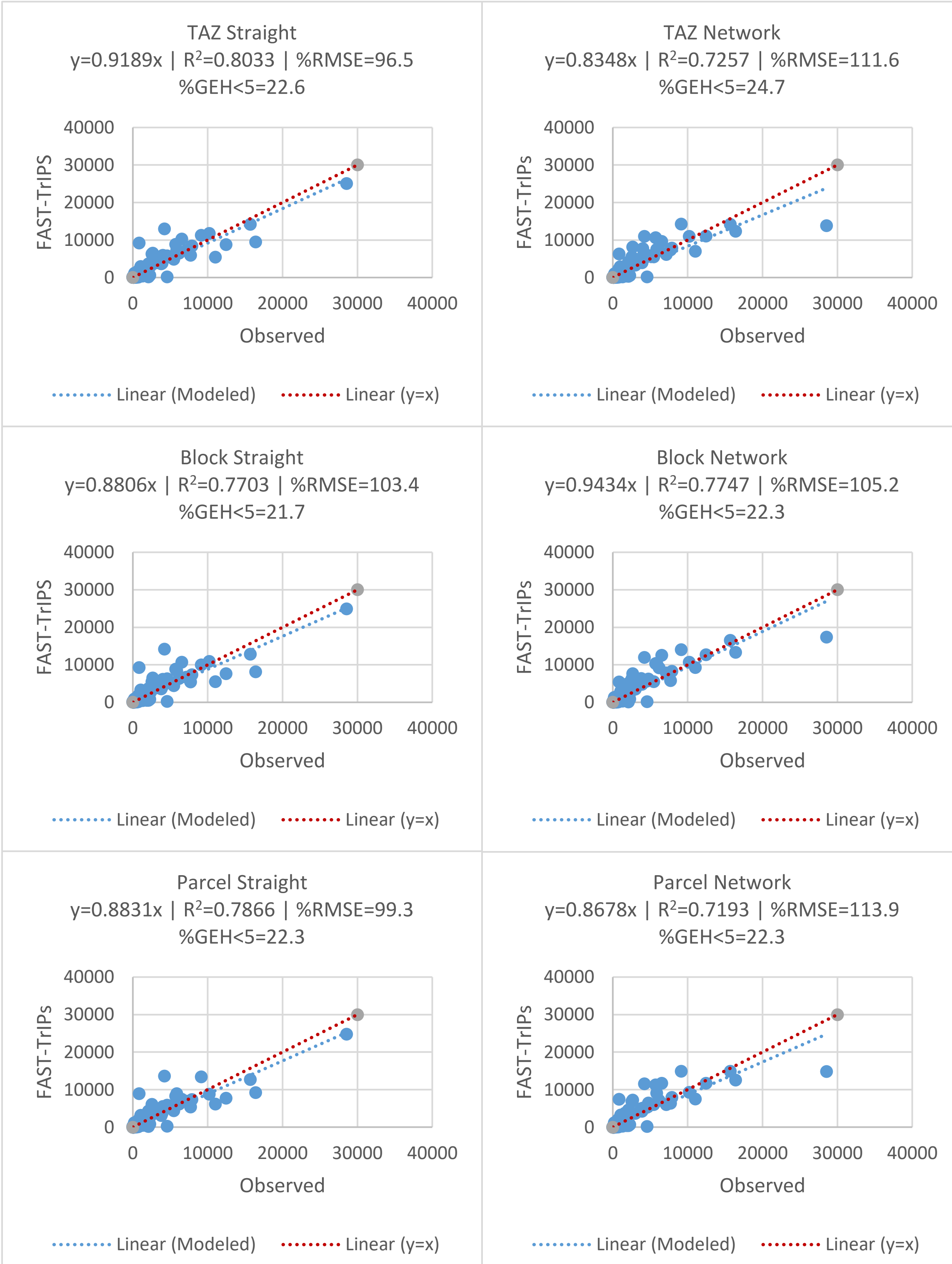

Figure 7: System Ridership by Route

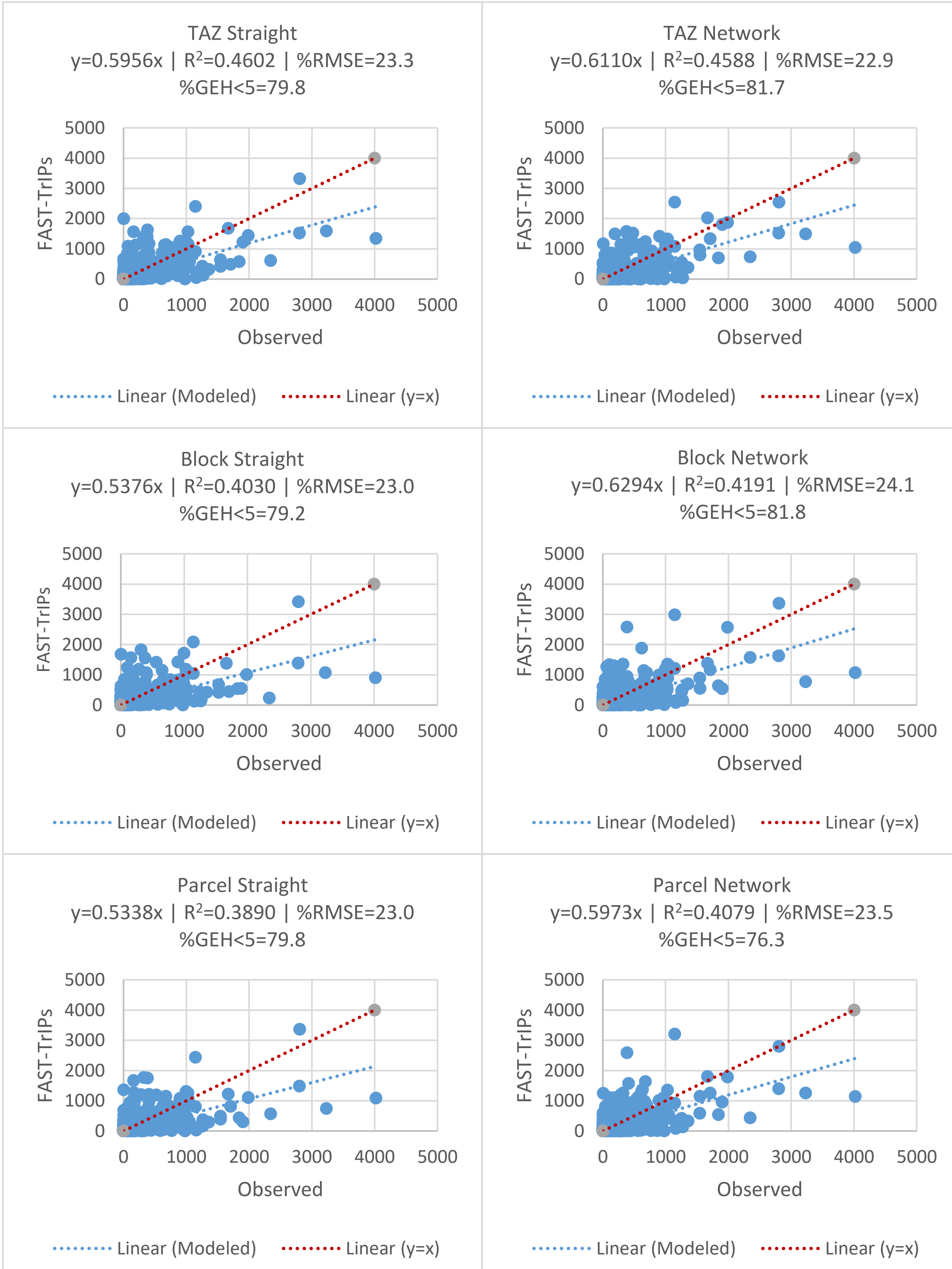

Figure 8: System Ridership by Stop

5.7 Route Level Analysis

On a route level, the block network scenario shows improvement. In this analysis, three high ridership routes, routes 2, 16, and 21, were used as case studies. Figure 14, Figure 15, and Figure 16 show improvements in the load profiles calculated at a stop level as compared to the 2013 APC data's load profile. There were some troubles with matching the 2011 network to the 2013 data, as discussed in Section 5.4, but these routes matched the 2011 data well. Routes 2 and 16 show significant improvement in the block network scenario. Improvements were not seen in the Route 21 scenario, but worse results aren't seen except for the parcel straight scenario. These routes show where this methodology can improve the model's assignment of passengers, but improvements aren't seen in every scenario. Great GEH measures are observed in these routes, as all fall above the 60% cutoff for transit assignment. To get another perspective, visualizations of the model outputs can be projected on a map. Figure 10 and Figure 11, along with Figure 12 and Figure 13 show comparisons between boardings, alightings, and load for TAZ Straight and Block Network scenarios for the local bus routes 2 and 16, respectively. It can be seen the block network scenario shows noticeable differences in boardings, alightings, and load when compared to the original scenario. This is important to note, as all three cases show the sensitivity to changes when only the walking links file is changed.

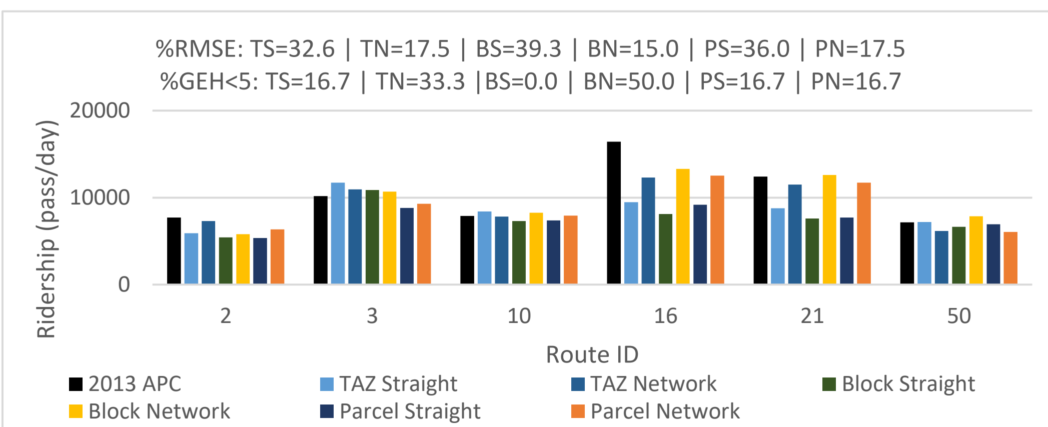

Figure 9: Total Ridership for High Ridership Routes

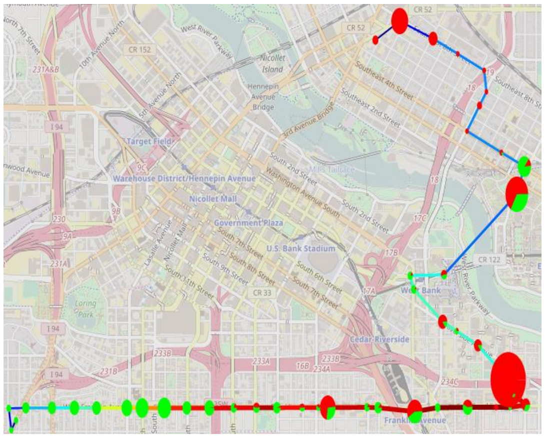

Figure 10: Route 2 Eastbound Morning Peak, TAZ Straight

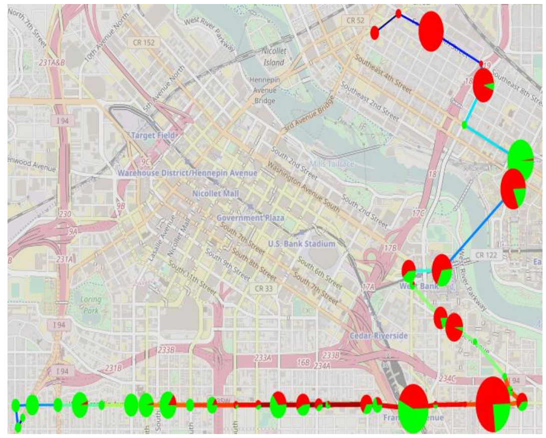

Figure 11: Route 2 Eastbound Morning Peak, Block Network

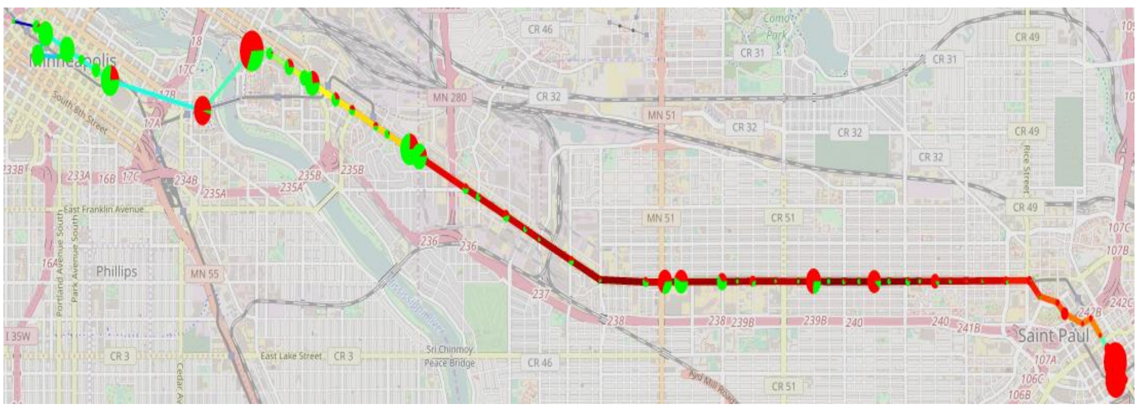

Figure 12: Route 16 Eastbound Morning Peak, TAZ Straight

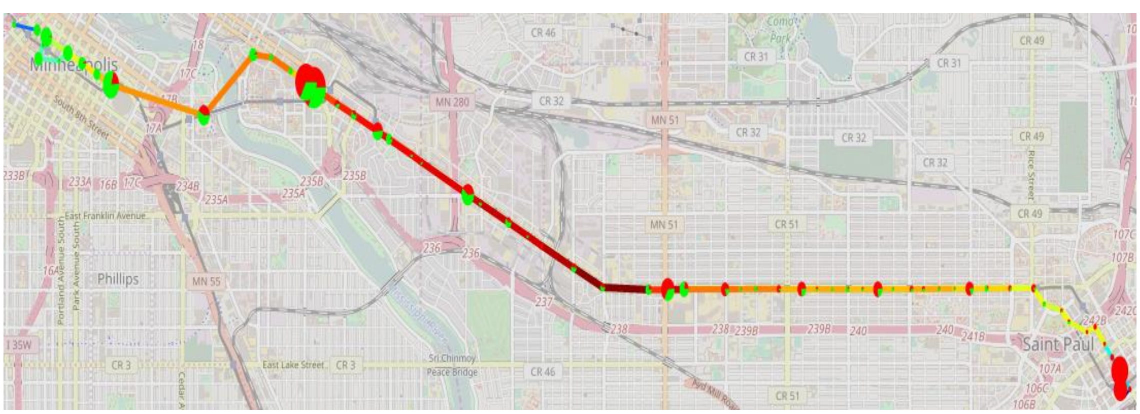

Figure 13: Route 16 Eastbound Morning Peak, Block Network

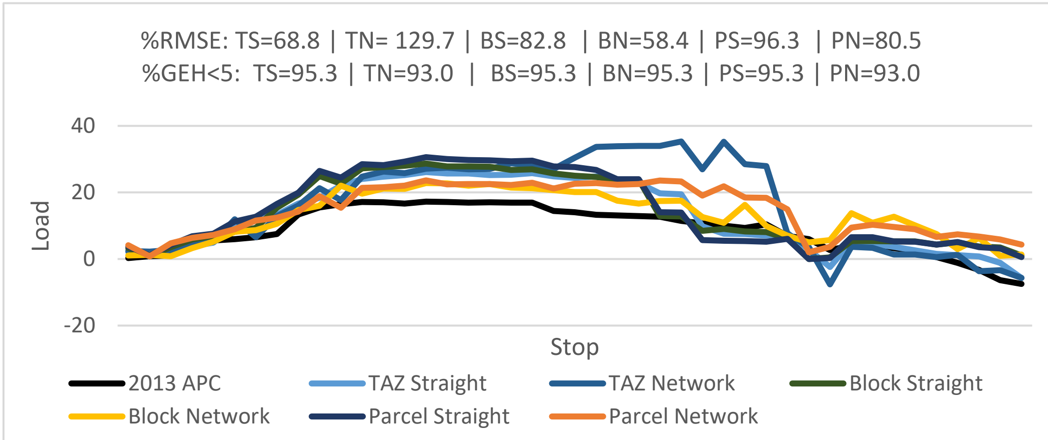

Figure 14: Average Daily Load for Local Bus Route 2

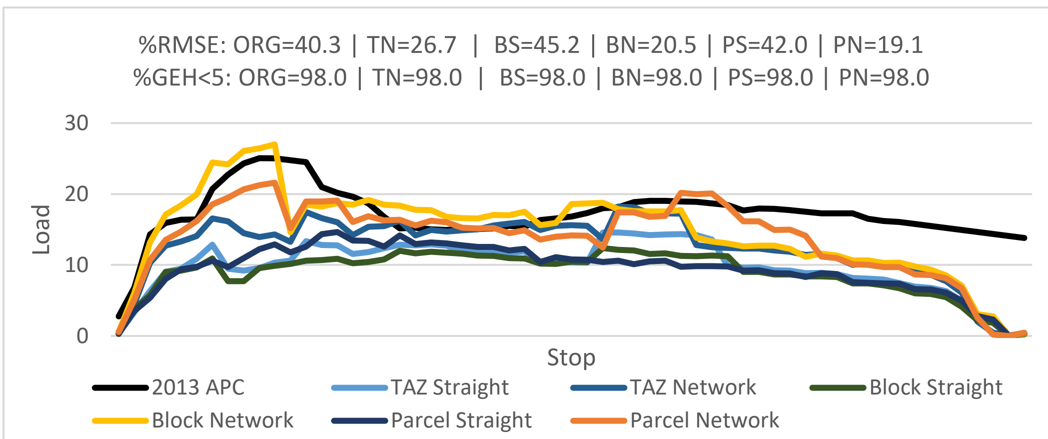

Figure 15: Average Daily Load for Local Bus Route 16

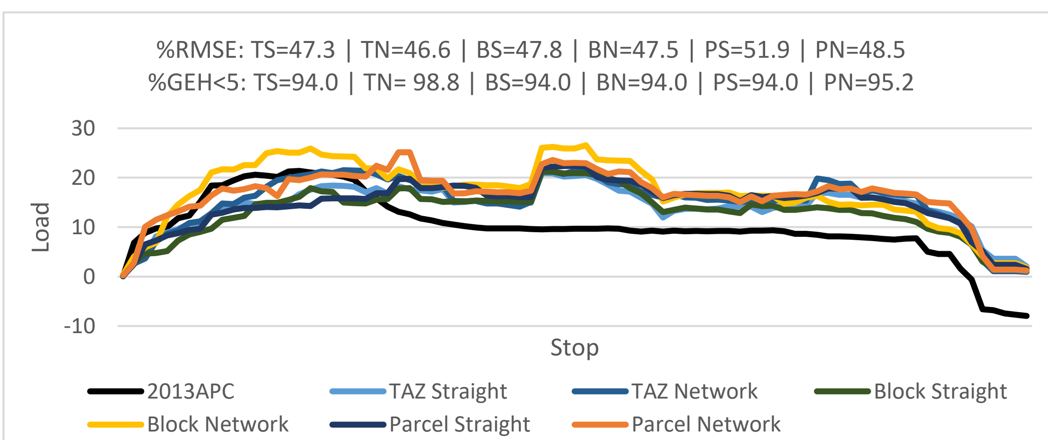

Figure 16: Average Daily Load for Local Bus Route 21

5.8 Neighborhood Level Analysis

To see how the methodology worked on a neighborhood level, four cases were studied: a residential area in the Northeast Minneapolis Arts District, the University of Minnesota East Bank Campus, a suburb of the Twin Cities (St Louis Park, MN), and the Uptown neighborhood in Minneapolis, MN. The university campus does not show a great fit when looking at the %RMSE and GEH measures. However, in the St Louis Park, Northeast Minneapolis Arts District, and Uptown scenarios, the availability of the population data show improvements of fit, particularly in the St Louis Park and Northeast Minneapolis Arts District. This shows that when consistent population and employment data (or other weighting measures representing transit demand) are available, improvements are seen.

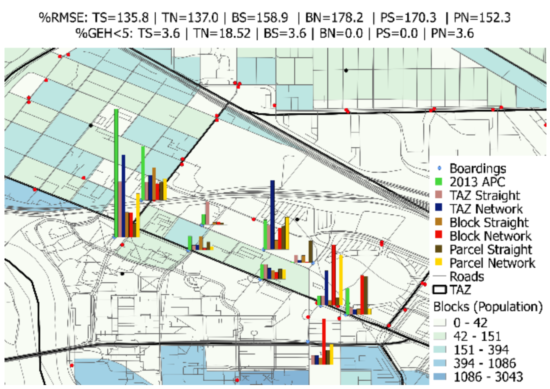

Figure 19: Boardings by Stop for University of Minnesota East Bank Campus

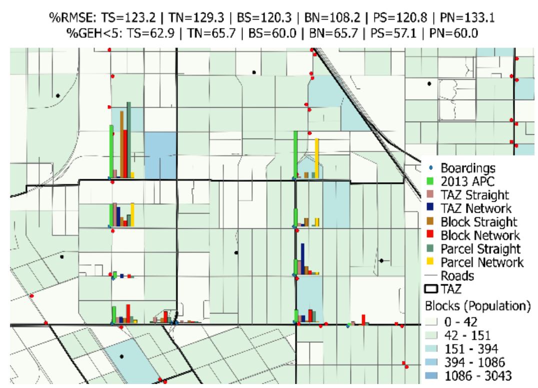

Figure 18: Boardings by Stop for North East Minneapolis Residential Neighborhood

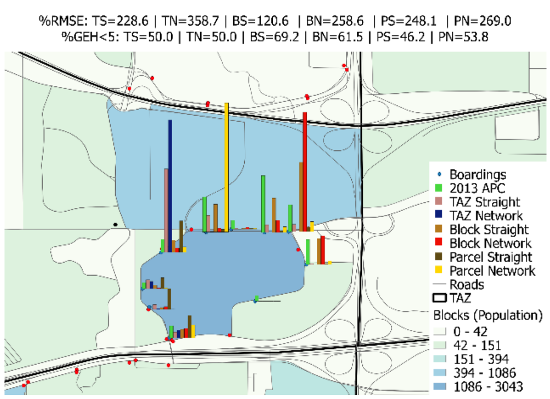

Figure 20: Boardings by Stop for Neighborhood in St Louis Park, MN

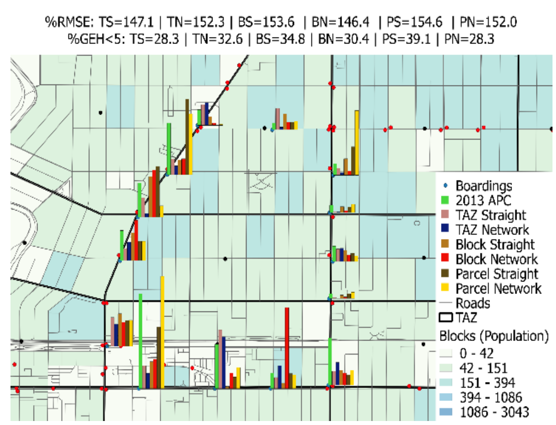

Figure 17: Boardings by Stop for Uptown, MN

5.9 Path Level Analysis

To determine if the model improved path prediction, analysis was done on the paths riders were taking from the three areas mentioned above as well as from the Como neighborhood near the university campus. The modeled paths were compared to the ridership from the 2010 Travel Behavior Inventory On Board Survey. Figure 22, Figure 23, and Figure 24 show the ridership on routes taken from St Louis park to downtown, Uptown to downtown, and the Como neighborhood to campus, respectively. Como hosts a large student population, and this scenario had the most data from the On Board Survey to compare against the scenarios.

When comparing the three scenarios, each one shows different scenarios performing the best. In Figure 23 the block and parcel straight scenarios show significant improvement, while in Figure 22 and Figure 24 they do not outperform the network scenarios. In Figure 22 and Figure 24 improvements in both the block and parcel network scenarios are seen. This shows just how detailed this transit assignment model can be, and it also shows how the different scenarios influence the level of analysis. This can also be seen in Figure 21, where multiple transit paths and the respective load is spatially presented for the options between the Como neighborhood and the downtown area. Generally, as the level of analysis gets lower, the smaller sized data units start to show the significance of this new methodology.

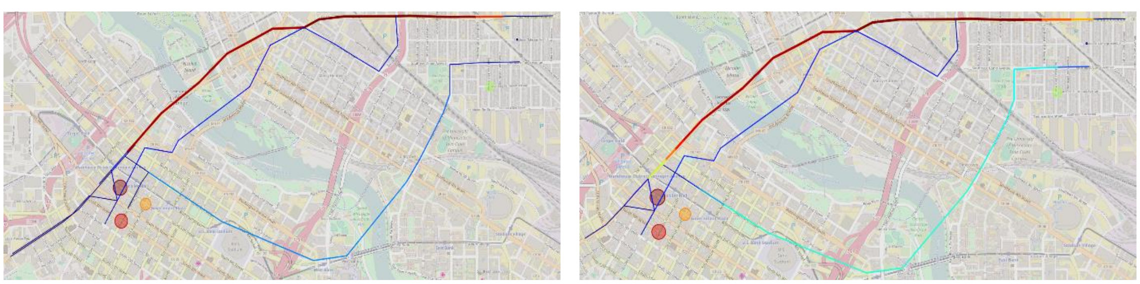

Figure 21: Visualized Paths for Como Area to Downtown. Left) TAZ Straight, Right) Block Network

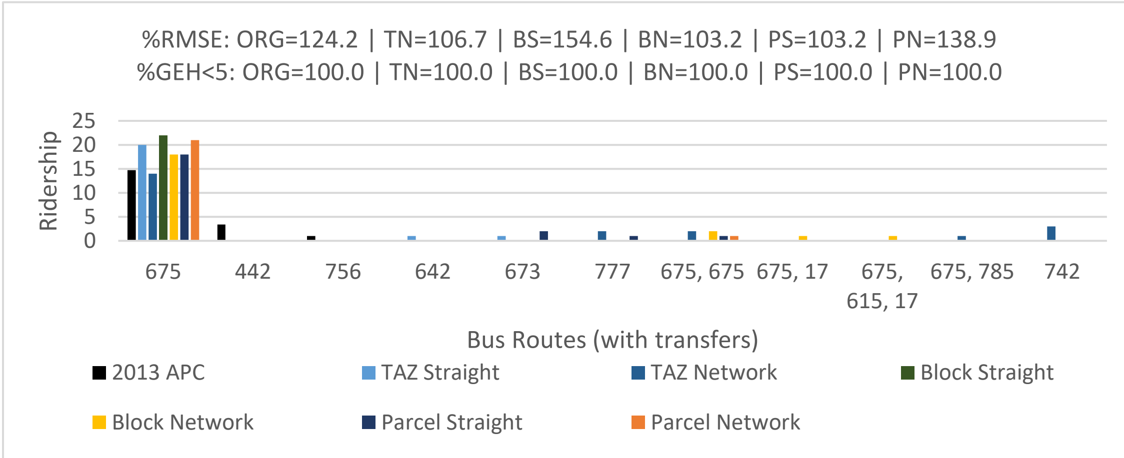

Figure 22: Transit Route Ridership from St Louis Park to Downtown

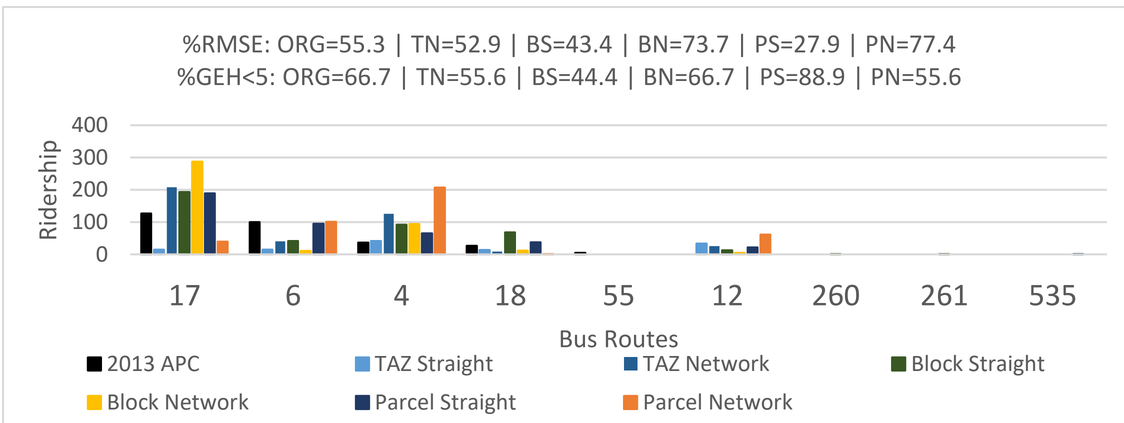

Figure 23: Transit Route Ridership from Uptown to Downtown

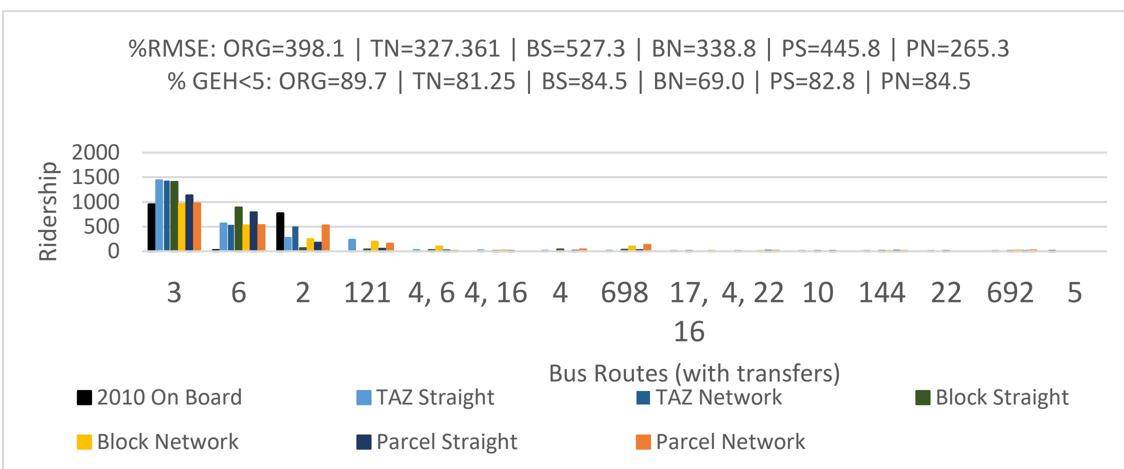

Figure 24: Transit Route Ridership from Como to University of Minneapolis (top 15/58 routes)

Chapter 6: Application in Park and Ride Access

6.1 Introduction

Park and ride access links are important as they are the second largest way people access transit after walking. As seen in Table 5 in Section 3.3, 79% of riders in the Twin Cities accessing transit by walking and 11% access transit by park and ride, according to the 2010 On Board Survey [29]. While walking is more difficult to calculate as it is much more sensitive to distance, park and ride poses a challenge as these riders often live in the outer suburbs of the city, meaning they still need to drive a distance to reach park and ride lots.

Due to the lower percentage of riders accessing transit through park and ride, this chapter will focus analysis on the improvements on transit assignment after including park and ride users. Due to the lack of improvements in the parcel level scenarios in the walking access links, these scenarios are ignored in this chapter as improvements are not expected when using driving links at that level. Instead, this chapter will only focus on the TAZ and block level straight line and network distances. All scenarios are weighted and calculated as they were in Chapter 5, and the same improvement measures are used.

6.2 Park and Ride Locations

Metro Transit has provided a dataset containing all of the park and ride locations on record [48]. This dataset includes information such as location, usage, capacity, and more in regards to the 239 park and ride and park and pool lots in the Twin Cities network. The lots included in this set are either closed, open, occasional, or proposed lots, as depicted in Table 11.

In order to prepare the data, the closed, occasional, and future lots were removed. Then out of the 152 remaining locations, 14 lots that were built after 2011 were removed to stay consistent with the transit network. Out of the remaining 138 lots, 47 park and pool lots were removed, as they are part of a carpooling service and are therefore not included in the assignment model. After removing lots that were not appropriate, 96 park and ride locations were left and used in this analysis.

Table 11: Park and Ride Network Characteristics

PR/PP Type	Count	Percentage
Open	152	64%
Closed	57	24%
Future	29	12%
Occasional	1	-
Total	**239**	**100%**

6.3 Limitations

Some notable limitations and assumptions were made when implementing park and ride access links into the transit assignment. Firstly, the locations of the park and ride lots were more spread out than the transit stops, and largely in the outer areas of the Twin Cities. This is a limitation as only assignment and land use data for the areas within the TAZ area are available. Therefore, many of the park and ride locations may be limited in their ridership due to the limit of the TAZ boundaries instead of the actual distances.

Secondly, a coarser network was used in this analysis than in the walking access calculations. Again, as the park and ride locations were more spread out, the detailed network used previously was not large enough to reach the park and ride locations. The new network was taken from the regional forecasting model provided by the Metropolitan Council of the Twin Cities. This network is used for traffic forecasting, and therefore doesn't contain the detail of small neighborhoods. Due to the coarseness of the network, lower levels of access in the network scenarios could be observed. The block scenarios also may not provide as much added detail as there are much fewer nodes to snap blocks onto the network, limiting the added detail of assignment from each block. Although the

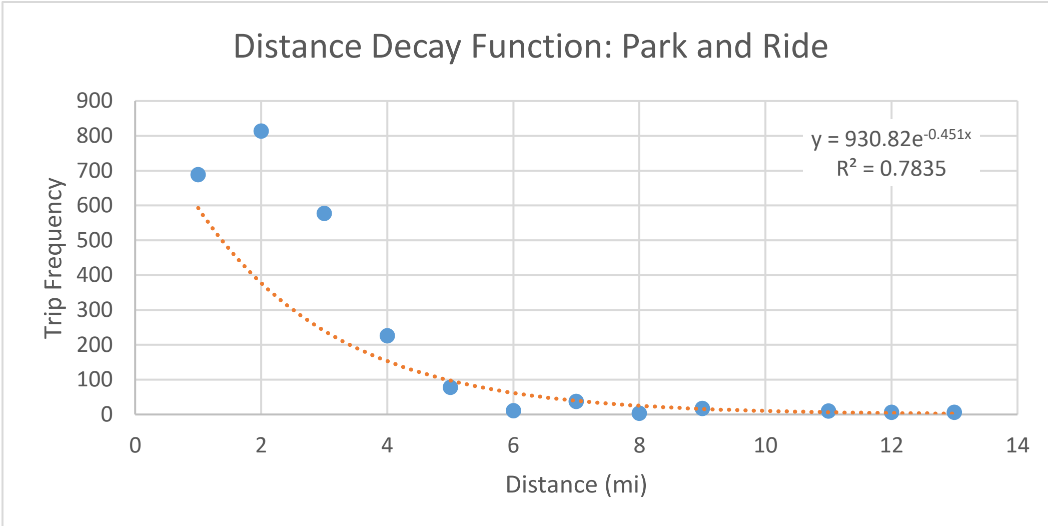

Figure 25: Distance Decay Function for Park and Ride Trips in the Twin Cities

added detail of the previous network would be preferred, this analysis focuses on driving access across larger distances, this coarse network is still appropriate.

Thirdly, it was assumed that riders would drive no longer than 10 miles (20 minutes at an assumed speed of 30 mph) to access park and ride lots. This threshold was chosen after calculating a distance decay function of how far people drove to park and ride lots based on the 2010 On Board Survey. Figure 25 depicts the decay function, and a sharp decrease in trips can be seen after 5 miles, and minimal trips after 10 miles.

Fourthly, in order to distinguish park and ride access and walking access, a new route choice parameter was used. The walking access link parameter, as seen in Table 8, was increased to be 10 times the value of in vehicle time. This is based on studies in other cities [49], as there currently are no route choice parameters estimated for park and ride access in the Twin Cities area. This was changed due to the different view park and ride users have of transit. These riders could drive to their destination, but view transit as a better option. This assumption can reasonably be made as the demand for this scenario is park and ride demand, not walking demand. This also influences the riders in the model to access park and ride locations nearest to their origin, whereas the current route choice parameter, 0.82, may influence these riders to access the park and ride closest to their destination as driving time is more desirable than transit time. This more accurately depicts the decisions of park and ride users.

Finally, a transit stop within 0.25 miles of a park and ride location was assumed as accessible. Often a park and ride location is located at a specific stop, but it doesn't need to be. Therefore, by allowing these stops to be accessed, riders who walk to a transit stop that the park and ride may not be intended for can be taken into account. Also, with a maximum walking time of 5 minutes, the time riders walk from their vehicle to the transit stop was not taken into account. As park and ride trips are typically longer than other types of transit trips and park and ride users tend to park close to their desired stop, this time is considered insignificant. With this assumption, the average park and ride lot had 6 accessible stops. It should be noted this assumption was not based on the 0.25-mile buffer distance often criticized in the literature. Instead, it was based on visual inspection of the distances from select park and ride facilities to reasonably accessible stops.

6.4 Access Links

The characteristics of the park and ride access links tend to be similar to the walking access links, as depicted in Table 11 and Figure 26. Initially, it may seem odd there are a similar amount of access links for a fraction of considered stops. However, due to the larger access distance, nearly all stops are accessible in each zone.

The average access trip distance is between 6 – 7 miles, which is reasonable based on the estimated distance decay function. As in the walking access calculations, decreases in the number of links for the network scenarios are seen. Again, this is to be expected given the difference in accessible area from a straight line distance as compared to a network distance of the same magnitude. However, it can be seen the average distance from a zone to a stop increases with added detail, with the highest average distance being in the block network scenario. This is strange as the network scenarios would be expected to have the lowest accessible distance. Unfortunately, this could be due to the coarser network used, as the snapping distance between the block and the network could be larger which would artificially increase the access area.

Table 12: Park and Ride Access Link Characteristics

Description	TAZ Straight	TAZ Network	Block Straight	Block Network
Number of Access Links	285,888	198,707	304,438	229,540
Zones with Accessible Stops	493	491	493	491
Number of Accessible Stops	2,442	2,278	2,496	2,384
Average Distance from Zone to Stop (mi)	6.77	6.93	7.14	7.52
Average Distance from Stop to Zone (mi)	6.57	6.84	6.83	7.23
Average Stops per Zone	117	87	122	96
Average Zones per Stop	580	405	618	467

Figure 26: Park and Ride Access Link Characteristics

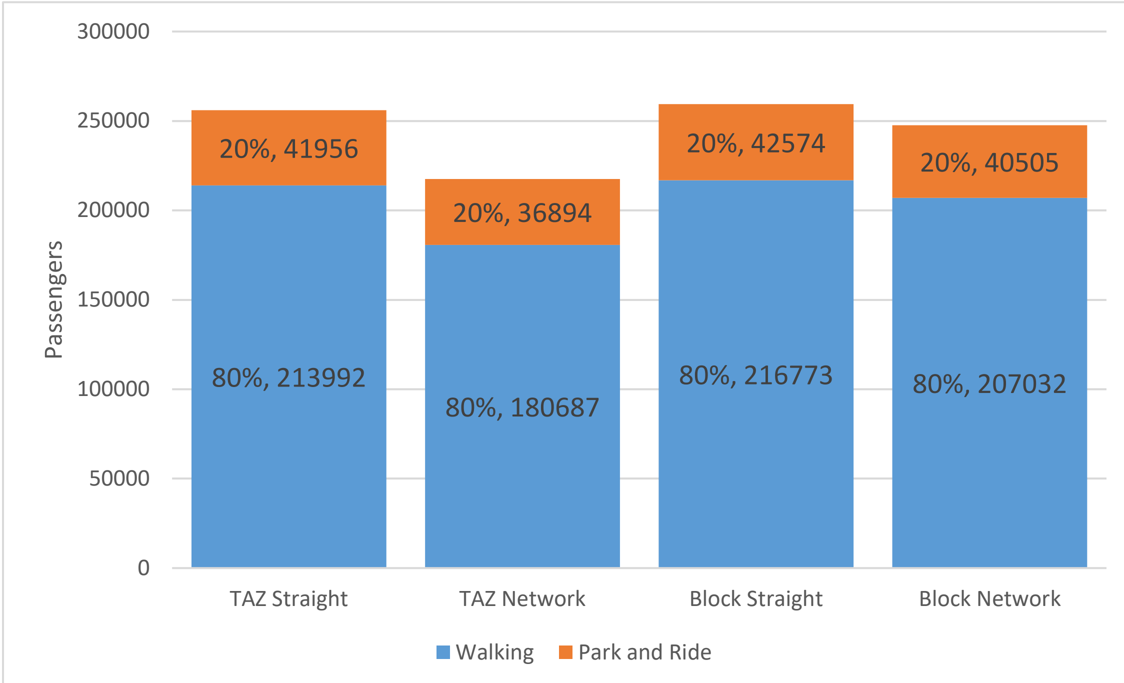

Figure 27: Assigned Passengers by Access Type

6.5 Assigned Passengers

In assessing the number of assigned passengers, the disparity between the straight line distances and the walking distances continue. Figure 27 depicts the total number of assigned passengers both due to walking links and park and ride links. From the On Board Survey the breakdown of walking access to park and ride access should be around 79% to 11%, as shown in Table 5. The results do show a similar proportion, especially as this analysis does not take drop off or bike access into account. Overall, the assignment is as expected, with a positive and close correlation between assigned passengers and number of access links.

6.6 System Level Analysis

On a system level, similar results of calculating ridership on a route and stop level can be seen, as depicted in Figure 28 and Figure 30 respectively. In analyzing these figures, it can be seen the park and ride access links do not improve ridership estimation. In fact,

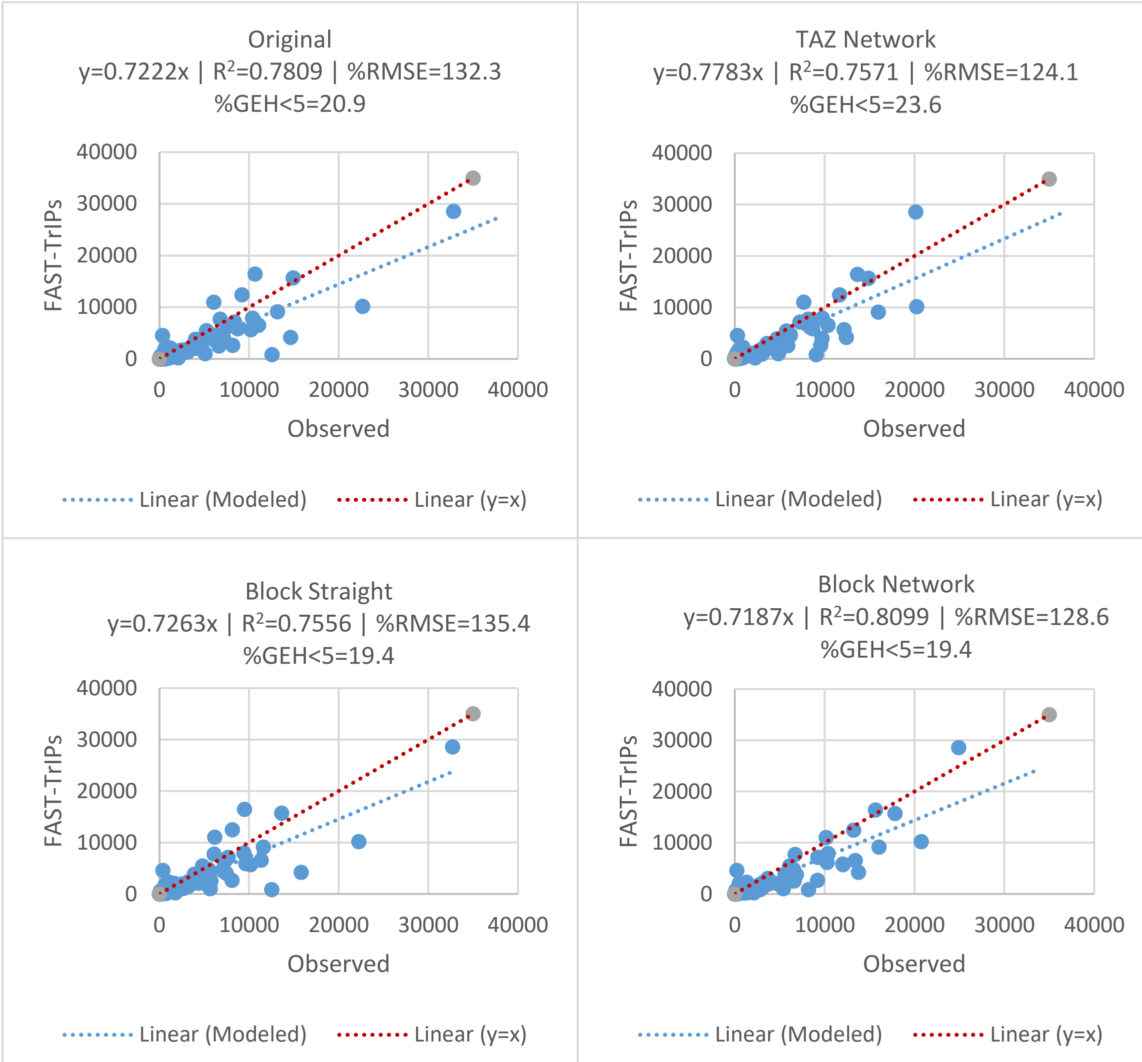

Figure 28: Ridership by Route (Walking and Park and Ride)

it decreases overall assignment accuracy. This could be due the many assumptions listed above, as well as the fact the access measurement is not park and ride specific. It does not implement a backtracking factor, instead it uses a much larger route choice parameter to influence park and ride choice.

However, the TAZ scenarios do slightly outperform the block scenarios. This is surprising after the walking access results, but may indicate the TAZ level is appropriate for driving links. This would make sense, as TAZs are designed specifically for traffic assignment, and are the zones used in regional traffic assignment models. This is also surprising as it may be expected that higher detailed scenarios, such as the block scenarios, would show improvement over the less detailed TAZ scenarios. However, due to the coarse network, these improvements may be less robust.

Figure 30 shows the system level boardings by stop. As with the walking access, accurate reproduction of observed values is not seen. However, in this case the block network scenario outperforms the TAZ scenarios, even though the stops seem more spread out. When comparing a stop to a route, the added detail of access distance should be more influential, which is noticed here. This is another example where the access link scenario's performance depends on the analysis required.

This stop access can also be assessed spatially, as depicted in Figure 31. These figures show park and ride usage based on scenario. When comparing the modeled usage to observed values, the modeled values often over predict usage. However, it can be seen that this over prediction is consistent across the scenarios. This is a good sign as there are not great spatial discrepancies in park and ride assignment. Therefore, the discrepancies could be due to additional factors such as demand as well as access link calculations.

6.7 Route Level Analysis

As the vast majority of park and ride facilities appeal to suburban populations, it should be no surprise that park and ride facilities are mainly located along express bus routes. Therefore, this analysis focuses on express routes instead of the routes previously used. Figure 29 shows the ridership of the 5 highest ridership express routes. These routes all have significant impact from park and ride services as they each have between 5 and 8 facilities in their routes. This makes them an appropriate example for this analysis.

In looking at how the park and ride users impact the highest ridership express routes, differences can be seen from the walking scenario. Mainly, the assignment is not as accurate, and there are more inconsistencies between scenarios. This would be expected based on the results from Figure 28 and Figure 30, but the estimates are more drastic. Specifically, ridership in Route 94 is drastically overestimated and almost doubled in the block straight scenario. Again, this is most likely due to the demand and general assignment calculation.

In looking at the ridership by route depicted in Figure 32, Figure 33, and Figure 34, general consistencies can be seen. These are radial routes, meaning they start in the suburbs and finish in the downtown area. These routes were specifically chosen based on the variety of attraction areas, with Route 250 from the Northeast suburbs, Route 850 from the Northwest suburbs, and Route 675 from the Western suburbs. In all three routes, spikes in the observed ridership can be seen. Specifically, in Routes 250 and 850, these spikes seem drastic. Increases in ridership should be expected as the route enters the more urban area, but The APC data shows a large increase in the number of boardings at specific park and ride lots on these routes. These figures represent the average daily load for these routes, and therefore it can be determined the access model fails to accurately represent this surge in boardings at this lot. For example, the stop where the Route 250 load increases has a predicted average boarding of over 150 passengers, whereas none of the modeled scenarios represent this. Inconsistencies between scenarios can still be seen, however they are mainly in the magnitude of load rather than the timing of boardings and alightings. This could indicate the inaccuracies are more influenced by demand factors than spatial accessibility factors. Overall, the TAZ network scenario tends to perform best.

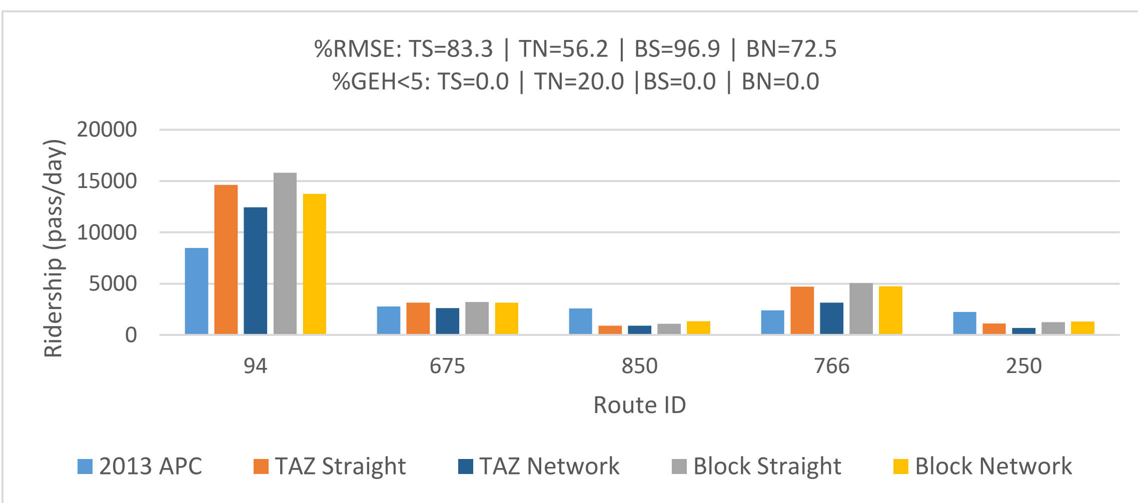

Figure 29: Total Ridership for Highest Ridership Express Routes

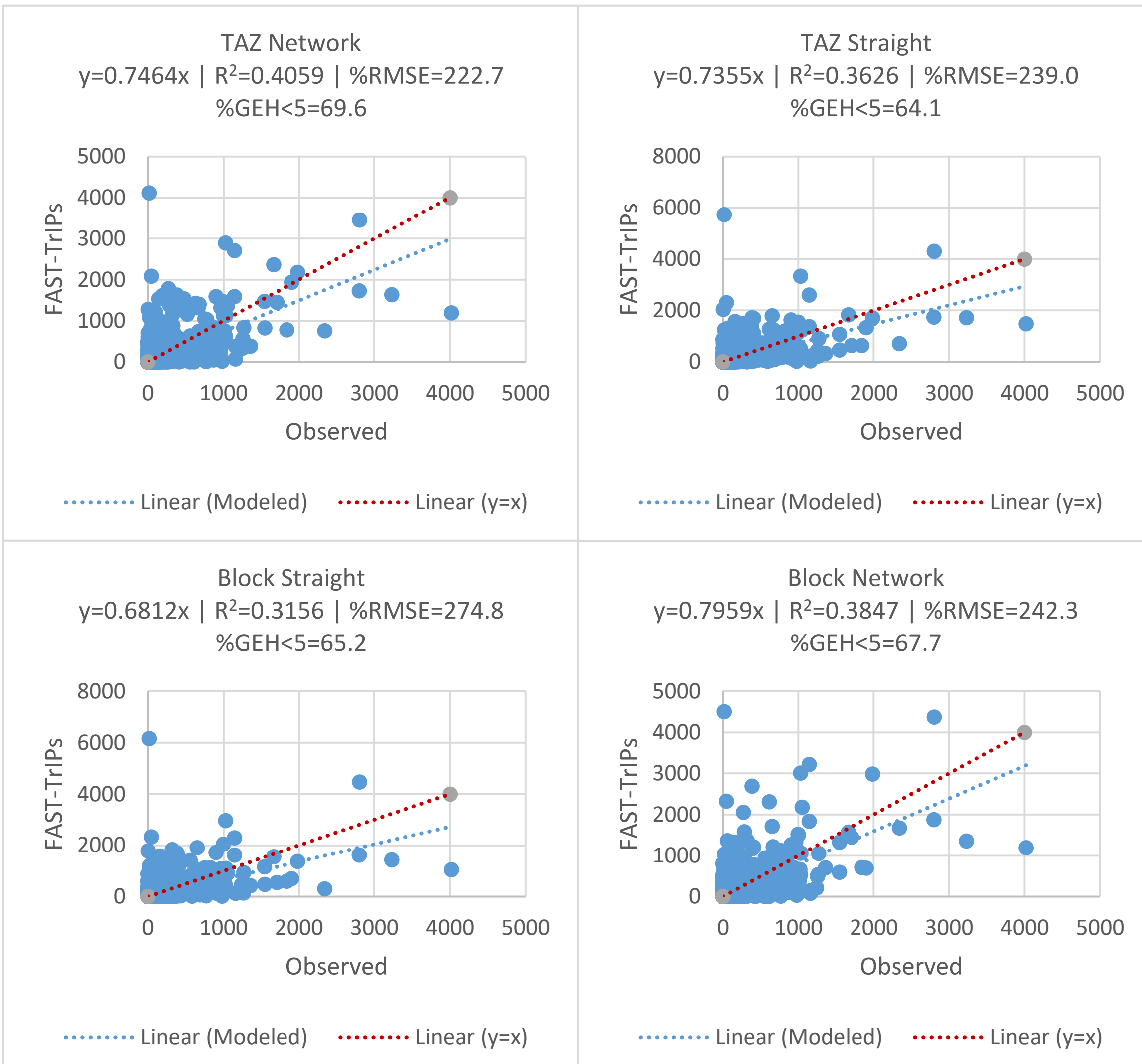

Figure 30: Ridership by Stop (Walking and Park and Ride)

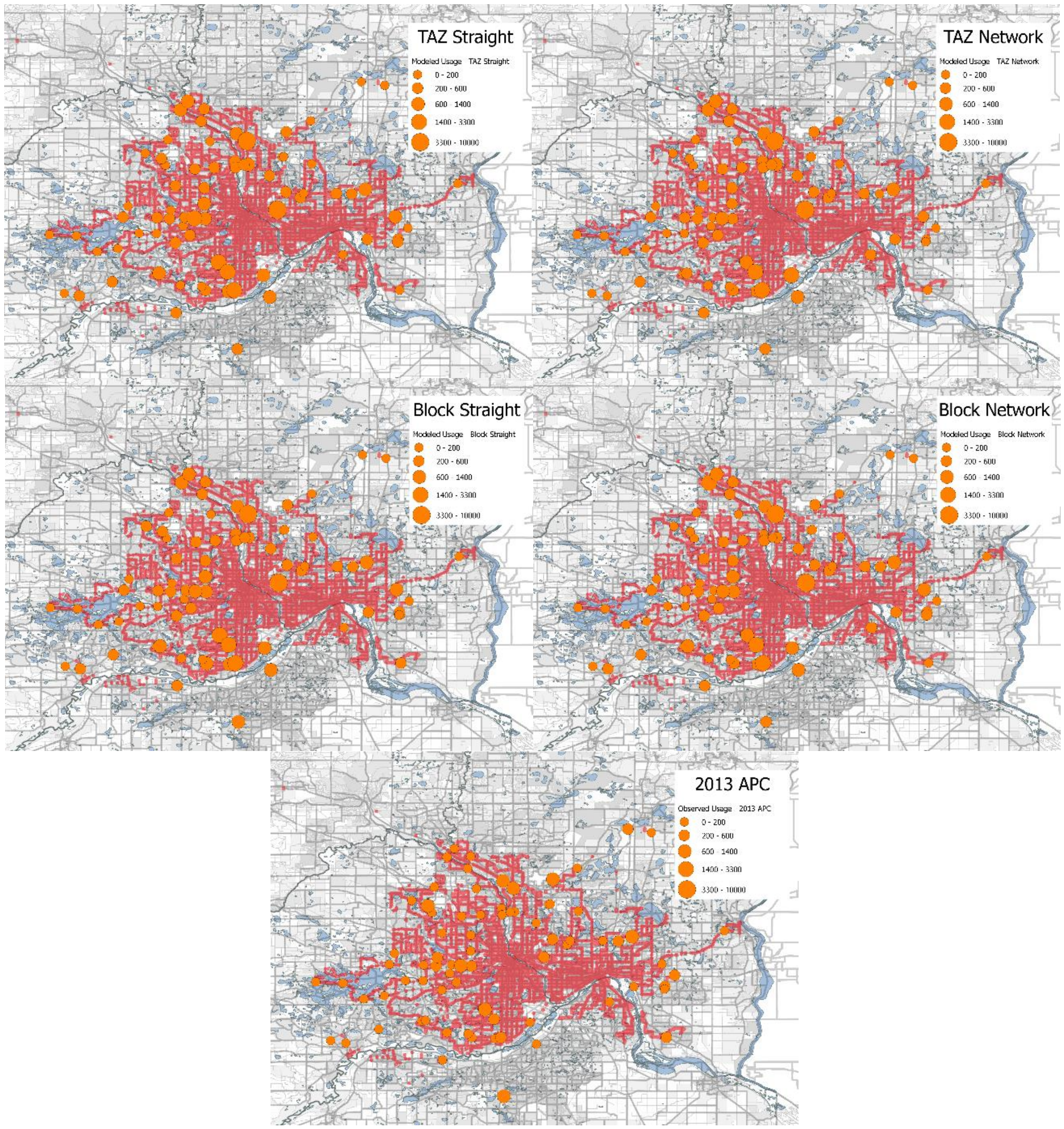

Figure 31: Park and Ride Usage

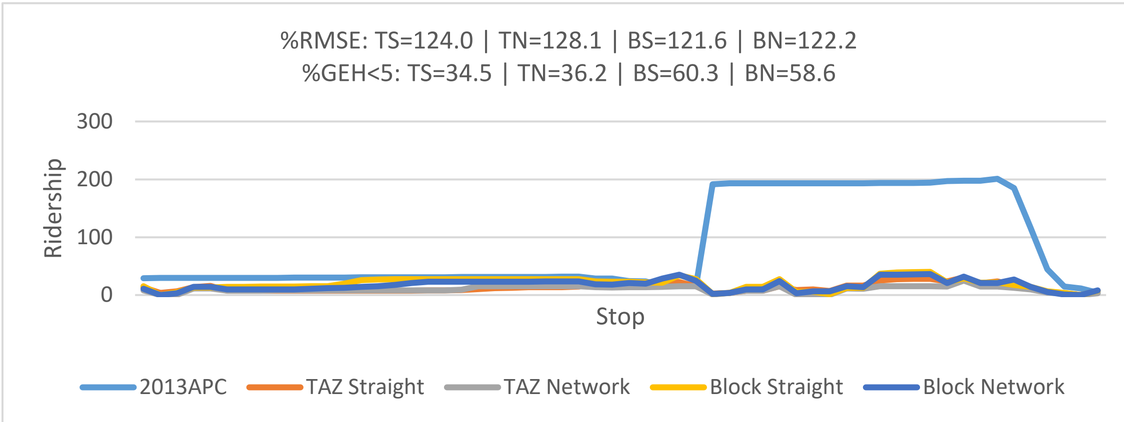

Figure 32: Average Daily Load for Express Bus Route 250

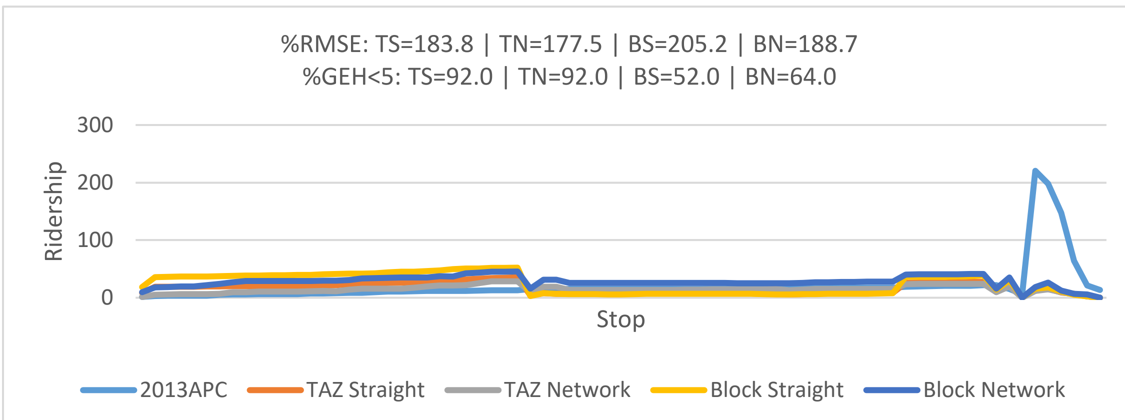

Figure 33: Average Daily Load for Express Bus Route 850

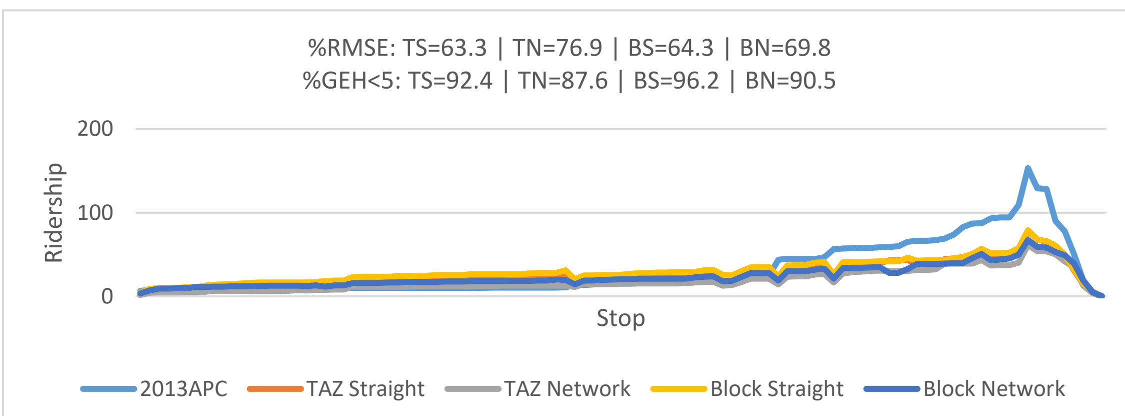

Figure 34: Average Daily Load for Express Bus Route 675

Chapter 7: Discussion and Conclusion

7.1 Discussion

Future research may include implementing this method using population, land use, or other data for aggregation on a parcel or block level, if this type of data is available. There was no improvement found when parcel data was aggregated by area, but this isn't to say this type of analysis wouldn't provide benefit in other cases. Also, a combination of data such as population, area, and land use could be combined to form a more holistic view of the demand from the parcel or block. Using a more up-to-date route choice model and sensitivity analysis to the logit dispersion factor in the proposed method could be done to improve accuracy.

Due to the disparity in the walking link accuracy due to location, a custom land use factor may be appropriate. That is, the blocks could be identified with a certain land use, and the best method of distance calculation could be determined. Therefore, TAZs that are comprised of more commercial or industrial land can utilize the straight line aggregation while TAZs with more residential area can utilize the network distance. This method could provide a more consistent improvement for the walking access link calculations by utilizing the best distance aggregation for the TAZ's land use.

Future analysis could also apply methods to handle bias in the %RMSE calculation and develop ways to determine significance when comparing ridership. The %RMSE measure is inherently biased as if the number of assigned passengers is greater than the number of observed passengers, the %RMSE will never show perfect results. Also, the observed dataset (2013 APC) shows ridership for unlinked trips while the modeled results are in the form of linked trips. Therefore, if scenarios have a higher transfer rate than the observed ridership, the modeled ridership will be artificially high. As this research mainly focuses on the redistribution of these passengers, the error terms could account for this bias and focus more on the distribution of passengers throughout the network. Also, in much of the analysis it was difficult to determine if an improvement was statistically significant.

By implementing a statistical test along the lines of a t-test or an F-test, the statistical significance of the scenarios provided in this research could be determined.

7.2 Conclusion

In this study, a new method to model transit access links is presented for application in transit assignment. This method provides higher levels of detail by using smaller land units (e.g. blocks, parcels) without sacrificing computational complexity of the model. Five different scenarios with varying aggregation measures, land units, and distance calculations were tested and compared against the current modeling scenario. For walking access, the scenario using block level population and employment data along with network distance showed the most consistent improvement across many levels of analysis, while the scenarios that didn't use network distance or didn't use data-driven aggregation weighting didn't show consistent improvements. Therefore, it is suggested to implement this data weighted method for walking access links if consistent and homogenous population data and network data is present and lower level analysis, such as path or route, is required.

As for park and ride locations, the more detailed block scenarios didn’t show consistent improvements. The more general TAZ scale scenarios performed better than the lower level block scenarios. This could be due to the much larger access distances where the TAZ’s size is more appropriate. Therefore, the TAZ network scenario may be used for park and ride access.

References

[1] F. Zhao, L.-F. Choq, M.-T. Li, A. Gan and I. Ubaka, "Forecasting Transit Walk Accessibility: A Regression Model Alternative to the Buffer Method," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1835, no. 1, pp. 34-41, 2003.

[2] T. Lei, C. Yali and G. Konstadinos, "Opportunity-Based Dynamic Transit Accessibility in Southern California," *Transportation Research Record: The Journal of Transportation Research Board,* vol. 2276, no. 1, pp. 26-37, 2012.

[3] S. A. Mamun, N. E. Lownes, J. P. Osleeb and K. Bertolaccini, "A Method to Define Public Transit Opportunity Space," *Journal of Transport Geography,* vol. 28, pp. 144-154, 2013.

[4] P. Anderson, A. Owen and D. Levinson, "The Time Between: Continuouslt-defined accessibility functions for schedule-based transportation systems," in *1st European Symposium on Quantitative Methods in Transportation Systems*, 2012.

[5] S. O'Sullivan and J. Morrall, "Walking Distances to and from Light-Rail Transit Stations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1538, pp. 19-26, 1996.

[6] Metro Transit, "GTFS Data," 2016. [Online]. Available: https://developers.google.com/transit/gtfs/.

[7] S. Biba, K. Curtin and G. Manca, "A New Method for Determining the Population with Walking Access to Transit," *International Journal of Geographic Information Science*, vol. 24, no. 3, pp. 347-364, 2010.

[8] Z.-R. Peng, K. J. Dueker, J. Strathman and J. Hopper, "A Simultaneous Route-Level Transit Patronage Model: Demand, Supply, and Inter-route Relationship," *Transportation,* vol. 24, pp. 159-181, 1997.

[9] J. L. E. Andersen and A. Landex, "GIS-based Approaches to Catchment Area Analyses of Mass Transit," in *ESRI International User Conference*, 2009.

[10] T. J. Kimpel, K. J. Dueker and A. M. El-Genneidy, "Using GIS to Measure the Effect of Overlapping Service Areas on Passenger Boardings at Bus Stops," *Urban and Regional Information Systems Association Journal (URISA),* vol. 19, no. 1, pp. 5-11, 2007.

[11] W. Lam and J. Morrall, "Buss Passenger Walking Distances an Waiting Times: A Summer-Winter Comparison," in *RTAC Annual Conference Preprints,* 1981.

[12] A. Khani, B. Bustillos, H. Noh, Y.-C. Chiu and M. Hickman, "Modeling Transit and Intermodal Tours in a Dynamic Multimodal Network," in *Transportation Research Board 93rd Annual Meeting,* Washington D.C., 2014.

[13] H. S. Leninson and O. Brown-West, Estimating Bus Ridership, 1984.

[14] A. Tavassoli, M. Mesbah and M. Hickman, "Quantifying Error in Transit Assignment Using Smart Card Data in a Large-Scale Multimodal Transit Network," in *Transportation Research Board 94th Annual Meeting*, Washington, D.C., 2016.

[15] F. Zhao, "GIS Analysis of the Impact of Community Design on Transit Accessibility," in *Proceedings of the ASCE South Florida Section Annual Meeting*, 1998.

[16] S. Grava, Urban Transportation Systems: Choices for Communities, New York: McGraw-Hill, 2003.

[17] A. El-Geneidy, M. Grimsrud, R. Wasfi, P. Tetreault and J. Surprenant-Legault, "New Evidence on Walking Distances ato Transit Stops: Identifying Redundancies and Gaps using Variable Service Areas," *Transportation,* vol. 41, pp. 193-210, 2014.

[18] J. Larsen, "Beyond the Quarter Mile: Re-examining Travel Distances by Active Transportation," *Canadian Journal of Urban Research,* vol. 19, no. 1, pp. 70-88, 2010.

[19] Y. Yang and A. V. Diex-Roux, "Walking Distance by Trip Purpose and Population Subgroups," *American Journal of Preventative Medicine,* vol. 43, no. 1, pp. 11-19, 2012.

[20] H. Badland, M. White, G. MacAulay, S. Eagleson, S. Mavoa, C. Pettit and B. Giles-Corti, "Using Simple Agent-based Modeling to Inform and Enhance Neighborhood Walkability," *International Journal of Health Geographics,* vol. 12, no. 1, p. 58, 2013.

[21] M. W. Horner and A. T. Murray, "Spatial Representation and Scale Impacts in Transit Service Assessment," *Environment and Planning B: Planning and Design,* vol. 31, pp. 785-797, 2004.

[22] Y. Jiang, P. C. Zegras and S. Mehndiratta, "Walk the Line: Station Context, Corridor Type and Bus Rapid Transit Walk Access in Jinan, China," *Journal of Transport Geography,* vol. 20, pp. 1-14, 2012.

[23] A. W. Agrawal, M. Schlossberg and K. Irvin, "How Far, by Which Route and Why? A Spatial Analysis of Pedestrian Preference," *Journal of Urban Design,* vol. 13, no. 1, pp. 81-98, 2008.

[24] J. Gutierrez, O. D. Cardozo and J. C. Garcia-Palomares, "Transit Ridership Forecasting at a Station Level: An Approach Based on Distance-Decay Weighted Regression," *Journal of Transport Geography,* vol. 19, pp. 1081-1092, 2011.

[25] M. Kuby, A. Barranda and C. Upchruch, "Factor sInfluence Light-Rail Station Boardings in the United States," *Transportation Research Part A,* vol. 38, pp. 223-247, 2004.

[26] W. A. O'Neill, R. D. Ramsey and J. Chou, "Analysis of Transit Service Areas Using Geographic Information Systems," *Transportation Research Record,* vol. 1364, pp. 131-138, 1992.

[27] Z. Peng and K. J. Dueker, "Spatial Data Integration in Route-Level Transit Demand Modeling," *Journal of Urban and Regional Information System Association,* vol. 7, no. 1, pp. 26-37, 1995.

[28] S. Meek, S. Ison and M. Enoch, "Role of Bus-Based Park and Ride in the UK: A Temporal and Evaluative Review," *Transport Reviews,* vol. 28, pp. 781-803, 2008.

[29] Metro Transit, "Travel Behavior Inventroy (TBI) 2010 On Board Survey," 26 02 2013. [Online]. Available: ftp://ftp.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_metc/society_tbi_transit_onboard2010/metadata/metadata.html.

[30] M. Shirgaokar and E. Deakin, "Study of Park-and-Ride Facilities and Their Use in the San Fransisco Bay Area of California," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1927, no. 1, pp. 46-54, 2005.

[31] S. Hendricks and M. Outwater, "Demand Forecasting Model for Park-and-Ride Lots in King County, Washington," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1623, pp. 80-87, 1998.

[32] B. Farhan and A. T. Murray, "A GIS-Based Approach for Delineating Market Areas for Park and Ride Facilities," *Transactions in GIS,* vol. 9, no. 2, pp. 91-108, 2005.

[33] United States Census Bureau, "Longitudinal Employer-Household Dynamics," 2016. [Online]. Available: https://lehd.ces.census.gov/data/.

[34] Open Street Map (OSM), "Open Street Map," 2016. [Online]. Available: https://mapzen.com/data/metro-extracts/metro/minneapolis-saint-paul_minnesota/.

[35] Google Transit APIs, "GTFS Static Transit Overview," 12 07 2016. [Online]. Available: https://developers.google.com/transit/gtfs/reference/.

[36] Metro Transit, "Transit Stops Boardins and Alightings (Fall 2015)," Metro Transit, Minneapolis, 2016.

[37] S. Nguyen and S. Pallottino, "Equilibrium Traffic Assignmnet for Large Scale Transit Networks," *European Journal of Operational research,* vol. 37, pp. 176-186, 1988.

[38] A. Khani, "FAST-TrIPs," [Online]. Available: https://github.com/akhani/FAST-TrIPs.

[39] "Moving FAST-TrIPs from Research to Practice...a three-agency experience," 2015. [Online]. Available: http://fast-trips.mtc.ca.gov/.

[40] PB Consulting, Inc, "Twin Cities Regional Model: Development of the Twin Cities 2000 Mode Choice Model," Metropolitan Council of the Twin Cities, Minneapolis, 2004.

[41] QGIS, "QGIS," QGIS, [Online]. Available: http://www.qgis.org/en/site/index.html.

[42] Python Software Foundation, "Pythong.org," Python Software Foundation, [Online]. Available: https://www.python.org/.

[43] E. W. Dijkstra, "A Note on Two Problem in Connexion with Graphs," *Numerisch Mathematik*, vol. 1, pp. 269-271, 1959.

[44] Federal Highway Administration, "Clibration of Microsimulation Models," 2017. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/tat_vol3/sect5.htm.

[45] R. Dowling, A. Skabardonis, J. Halkias, G. McHale and G. Zammit, "Guidelines for Calibration of Microsimulation Models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1876, pp. 1-9, 2004.

[46] The Department of Transport, "Traffic Appraisal in Urban Areas," The Department of Transport, 1996.

[47] MWH, "COASTCONNECT: Caloundra to Maroochydore Transport and Traffic Summary - Technical Analysis," Department of Transport and Main Roads, 2004.

[48] Metro Transit, "Park and Ride Lots," 06 01 2017. [Online]. Available: ftp://ftp.gisdata.mn.gov/pub/gdrs/data/pub/us_mn_state_metc/trans_park_and_ride_lots/metadata/metadata.html. [Accessed 02 2017].

[49] CDM Smith, "CAMPO Travel Demand Model Validation Report," Capitol Area Metropolitan Planning Organization, Austin, 2014.